UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular I

RELACIÓN ESTRUCTURA-FUNCIÓN DEL SUPRESOR TUMORAL ING4.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Alicia Palacios Ortega

Bajo la dirección del doctor

Francisco José Blanco Gutiérrez

Madrid, 2010

- **ISBN: 978-84-693-2384-7**

Universidad Complutense de Madrid.
Facultad de Ciencias Químicas.
Departamento de Bioquímica y Biología Molecular I

Tesis Doctoral

RELACIÓN ESTRUCTURA-FUNCIÓN DEL SUPRESOR TUMORAL ING4

ALICIA PALACIOS ORTEGA

Madrid 2oo9

Tesis Dirigida por:
Dr. Francisco José Blanco Gutiérrez

Realizada en:
Grupo de Resonancia Magnética Nuclear.
Programa de Biología Estructural y Biocomputación.
Centro Nacional de Investigaciones Oncológicas, CNIO.
Madrid

Grupo de Resonancia Magnética Nuclear.
Unidad de Biología Estructural.
Centro de Investigación Cooperativa en Biociencias, CIC-bioGUNE.
Derio, Bizkaia

Agradecimientos

¡Por fin!, ha llegado el momento de escribir los agradecimientos…
Bueno, en primer lugar agradecer a Jesus y Pilar, por ayudarme a dar los primeros pasos en esto y a Eli, Barbi, Jana y Almu por animarme a seguir intentando hacer lo que me gusta tras mi intento fallido en el L4/L7.
A Paco, mi director de tesis tengo mil cosas que agradecerle, en primer lugar su paciencia, su afán por enseñar, su facilidad para ilusionarme y a veces "liarme" con nuevos proyectos, su trato de igual a igual, su cariño (el que sabemos que nos tiene a pesar de su seriedad…) y lo más importante, su capacidad para escuchar, dialogar, discutir (a veces demasiado) y compartir nuestras ideas, muchas gracias Paco.
Como no, quiero agradecer a todos los compañeros del CNIO, a Ramón, Jesús, Eva, Simone, Dani T. y Elenita, por todos los buenos ratos, gracias. Especialmente, agradecer a Susana, por su gran amistad y ayuda, a Tahl, por su amistad y por tantas risas, bromas y conversaciones científicas, a Richard, mi eterno compañero de habitación, por su gran paciencia, compañía y amistad durante este año bilbaíno. A Pascal, por ser mi jefe y amigo. A Dani Padro y David Pantoja, que decir de vosotros, mis grandes compañeros, amigos y consejeros, gracias por ayudarme siempre.
De mi fase en el CNIO tampoco podré olvidar a nuestros compañeros, los cristalografos y especialmente a Mariajo. Agradecer también a Ines y Guillermo por su ayuda con los cristales. Y, por supuesto, a nuestros estudiantes: Bruno, mi gran amiga Irene y la alocada Lore, gracias. Y a ti Nuria, no se dónde meterte, gracias por estar siempre ahí.
El último año de la tesis lo hemos pasado en Bilbao, en el Biogune, de allí agradecer a todos los que nos acogieron, e hicieron posible que mi estancia en bilbo fuera una bonita experiencia: Pablo, Bea, Patri, Xavi, Danel, Ana, María, Borja, David… y a nuestras nuevas compañeras y amigas Nekane, Caterina y Maider, ekerrik asko y grazie tante.
De nuestras colaboraciones, me gustaría agradecer a toda la gente de granada (Jose Manuel, Manu, Irene ..), a los chicos del iib, especialmente a Alberto e Ignacio, y a Pau, nuestro último compañero con los INGs, gracias a todos.
Por último quería nombrar a toda esa gente que ha estado conmigo fuera del labo todos estos años, a mi madre y familia por estar ahí, a mis sobrinos que hacen que los malos días se conviertan en maravillosos, a mis amigas por serlo y, por supuesto, a quien me hace feliz cada día, Joni, por ayudarme con la tesis, por escucharme, comprenderme y aguantarme, gracias.
En definitiva gracias a todos los que estos años me han hecho sonreir..

ABREVIATURAS

ACF-1	Factor de complementación Apobec-1
ADN	Ácido desoxirribonucleico
ARN	Äcido ribonucleico
ASA	Área de superficie accesible
ATP	Adenosina-5´-trifosfato
BiFC	Complementación de fluorescencia bimolecular
BPTF	*Bromodomain PHD finger transcription factor*
CSP	Perturbación del desplazamiento químico
CD	*Circular dichroism*, dicroismo circular
CFP	Proteína fluorescente cián
CHD1	*Chromo-ATPase/helicase-DNA binding protein*
D1	Fragmento de ING4 (1-118)
D2	Fragmento de ING4 (119-187)
D3	Fragmento de ING4 (188-249)
D1D2	Fragmento de ING4 (1-188)
D2D3	Fragmento de ING4 (119-249)
DMSO	Dimetilsulfóxido
DSS	*Sodium 2,2-dimethylsilapentane-5-sulphonate*
DTT	Ditiotreitol
GST	Glutation S-transferasa
GTA	Glutaraldehido
H2A/2B/3/4	Histona 2A, 2B, 3 y 4
H3K4me3	Histona 3 trimetilada en lisina 4
$H3_{15}$K4me3	Péptido de 15 residuos de la histona 3 trimetilada en K4
HxAyBza	Histona x, modificada con z grupos (B: ac, acetilo me, metilo, ph, fosfato) en el residuo A en posición y. Cuando la modificación es asimetrica se coloca una "a" al final
HAT	Acetiltransferasa de histonas
HDAC	Desacetilasa de histonas
hEaf6	Proteína ortóloga humana del factor 6 asociado a Esa-1
HBO1	Acetiltransferasa de histonas de unión al complejo del origen de replicación I
HIF-1α	Factor de transcripción inducible por hipoxia
HMT	Metiltransferasa de histona
HNSCC	Carcinoma celular escamoso de cabeza y cuello
HP1	Proteína de la heterocromatina 1
HPH-2	Enzima prolil-hidroxilasa HIF
HRP	Peróxidasa de rábano picante
HSQC	*Heteronuclear single quantum coherence spectroscopy,* espectroscopía heteronuclear de cuanto único, en coherencia.
IL-6	Interleucina 6
IL-8	Interleucina 8
ING	INhibitor of Growth
ING4v	Variante de transcripción del gen ING4
IPTG	Isopropil β-D-1-tiogalactopiranósido
ITC	Calorimetría isoterma de titulación
LB	Medio de crecimiento de células Luria Bertani
LZL	*Leucine-Zipper-Like*, motivo tipo cremallera de leucinas
MCS	Sitio de clonación múltiple
MLL-1	*Mixed-lineage leukemia histone methyltransferase*, metiltransferasa de histonas de leucemia de linaje mixto
MOZ/MORF	Dominio de dedos de Zinc de leucemia monolítica y factor relacionado con él
MW	Peso molecular
NER	Reparación de nucleotidos por excisión
NF-κB	Factor nuclear *k*B
NLS	Secuencia de localización nuclear
NOE	*Nuclear overhauser effect*, efecto nuclear Overhauser
NOESY	*Nuclear overhauser enhancement spectroscopy*, espectroscopía de aumento del efecto nuclear Overhauser

NTS	Señal de direccionamiento al nucleolo
NURF	*Nucleosome remodelling factor*, factor de reestructuración nucleosómico
p14ARF	Producto de marco de lectura alternativo del gen CDKN2A
OD	Densidad óptica
OPN	Proteína osteopontina
PAGE	Electroforesis en gel de poliacrilamida
PBR	Región polibásica
PBS	Tampon fosfato salino
PCNA	Antígeno nuclear de proliferación celular
PCR	Dominio potencial de regulación de cromatina
PDB	*Protein Data Bank*
PDIM	Motivo de interacción dependiente de fosforilación
PHD	Homeo dominio de plantas, dominio de dedos de zinc
PI5P	D-myo-fosfatidilinositol 5- fosfato
pRb	Proteína del retinoblastoma
PtdInsPs	Fosfatidilinositol fosfatos
RKO	Línea celular derivada de cáncer colorrectal
RMN	Resonancia magnética nuclear
RMSD	*Root mean square deviation*, desviación media cuadrática
SAP30	Proteína asociada al complejo Sin3A
SAXS	Dispersión de rayos X a bajo ángulo
SDS	Dodecil sulfato sódico
Spp-1	Fosfoproteína secretada-1
TAF1	Factor asociado a TBP (proteínas de unión a la caja TATA)
TAP-tag	Purificación en tandem con columnas de afinidad
TAZ	Proteína con dedos de Zinc que funciona como adaptador en la transcripción
TGF-b	Factor de transformación del crecimiento B
TROSY	*Transverse relaxation optimized spectroscopy,* espectroscopía optimizada de relajación transversal
TSG	Gen supresor tumoral
UV	Ultravioleta
VFP	Proteína fluorescente Venus
VC	Venus (155-238)
VN	Venus (1-173)
WDR5	WD repeat-containing protein 5, proteína 5 que contiene dominios WD repetidos
WSTF	Factor de transcripción del síndrome Williams
WT	*Wild type*, forma salvaje
YFP	Proteína fluorescente amarilla
ZZ	Dominio putativo de unión a uno o más zinc

INDICE

INTRODUCCIÓN

Cáncer y genes supresores de tumores

El cáncer es una enfermedad en la que una célula o grupo de células adquieren la capacidad de dividirse sin control formando tumores que invaden los tejidos adyacentes, pudiendo llegar a diseminarse por el resto del cuerpo y provocar tumores secundarios (metástasis). Las causas y mecanismos de transformación de una célula normal en cancerosa son diversas y es necesario el concurso de varias de ellas para que la célula escape a los mecanismos de control del crecimiento. Desde un punto de vista genético destacan dos grupos de genes cuya alteración puede provocar o permitir el desarrollo del cáncer: los oncogenes y los genes supresores de tumores.

Los oncogenes provocan cáncer cuando son activados. Suelen ser genes dominantes ya que su efecto cancerígeno se debe a una ganancia funcional y la alteración de uno sólo de sus alelos es suficiente para su activación. En su versión inalterada (proto-oncogenes) tienen funciones muy diversas.

Los genes supresores de tumores son generalmente recesivos y es necesario que los dos alelos estén alterados para desactivarlos y que pueda progresar el tumor. Los supresores de tumores regulan directa o indirectamente el ciclo, proliferación y adhesión celulares, así como la reparación de daños en el ADN, transducción de señales, transcripción génica, e inducción de apoptosis (muerte celular programada).

La primera evidencia de la existencia de genes capaces de suprimir el crecimiento tumoral se obtuvo en 1969 (1) al observarse que tras la fusión de células cancerosas con ciertas células no cancerosas se perdían las propiedades tumorales de las primeras. De esto se dedujo que las células no cancerosas debían poseer factores de control que estaban ausentes en las cancerosas. El primer supresor tumoral identificado, la proteína del retinoblastoma (pRb), interacciona con ciertos factores de transcripción inactivándolos, lo cual inhibe la progresión del ciclo celular. Un papel central en muchos de los mecanismos de supresión tumoral lo ocupa el factor de transcripción p53, que está alterado en un 50 % de los tumores humanos.

1. La familia de supresores tumorales ING (*INhibitors of Growth*)

En 1996 se descubrió ING1, en un experimento diseñado para la identificación de supresores tumorales expresados diferencialmente en células del epitelio mamario en comparación con líneas celulares de cáncer de mama. El gen que codifica para ING1 fue clonado y sobreexpresado observándose su capacidad para inhibir el crecimiento celular, mientras que la expresión de un ARN antisentido de ING1 promovía la transformación celular, con lo cual se propuso que ING1 era un supresor tumoral (2).

Posteriormente, se han identificado otros cuatro genes homólogos (ING2-5) que componen la familia ING. Los cinco genes humanos se localizan en cinco diferentes cromosomas y, a excepción de ING3, en los extremos, cercanos a los telómeros. En diversos tipos de cáncer, se han encontrado mutaciones o deleciones de estos genes, y localización subcelular alterada de las proteínas ING (3), pero la causa más abundante de su inactivación parece ser la disminución o pérdida total de su expresión en células cancerosas (4). El hecho de que los genes de ING1,2, 4 y 5 estén localizados cerca de las regiones teloméricas sugiere la posibilidad de que la alteración de sus niveles de expresión observada en muchos tipos de cáncer se deba a la erosión del telómero que tiene lugar con el envejecimiento celular y durante la generación del tumor (5).

De los genes ING1, 3 y 4 se han descrito diferentes variantes postranscripcionales que comparten algunas funciones pero difieren en otras[1]. Se ha sugerido que las diferencias funcionales entre dichas variantes y el balance de sus niveles de expresión, podrían tener un papel importante en la generación del tumor (4). Las proteínas ING están implicadas, directa o indirectamente, en procesos muy diversos como reparación del ADN (6, 7), remodelación de la cromatina (8-11), parada del ciclo celular, inhibición de la proliferación celular (2) y la angiogénesis (12), e inducción de senescencia celular (13, 14) y apoptosis (15), (figura 1).

Se han descrito homólogos a las proteínas humanas en otros vertebrados, y también en invertebrados, plantas y hongos, todos ellos organismos eucarióticos. El análisis filogenético de las secuencias de la familia ING, indica que se pueden agrupar en tres subfamilias: ING1-2, ING4-5, e ING3, más distante evolutivamente de las dos primeras (5).

Las proteínas ING se localizan predominantemente en el núcleo, concentrándose durante algunos procesos en el nucleolo (donde se transcribe el ADN produciendo ARN ribosomal, y tienen lugar ciertos procesos de reparación de ADN). Esto ocurre en el caso de ING1 cuando aumentan los niveles de p14ARF (supresor tumoral implicado en la regulación del ciclo celular) y por efecto de la radiación ultravioleta (6, 16-18). También se ha observado la concentración de ING1 en la cara interna de la membrana nuclear (lámina nuclear) interaccionando con la proteína lámina A (19). Sin embargo, se han descrito también poblaciones minoritarias de algunas de las proteínas ING en el citoplasma. Así, una subpoblación de ING1 se localiza en el citoplasma cuando se sobreexpresa la proteína 14-3-3 (ocurre con todas las proteínas de la familia 14-3-3)(20). También se ha descrito la localización de ING4 en el citoplasma, concretamente en los lamelipodios, actuando como regulador de la propagación y migración celular a través de su interacción con la proteína citoplasmática liprina α1 (21).

1.1 Motivos de secuencia de las proteínas ING y su implicación funcional

El alineamiento de las secuencias de aminoácidos de las proteínas ING revela la existencia de regiones conservadas así como de ciertos motivos de secuencia y dominios estructurales en cada una de ellas. En general, las proteínas ING poseen una región N-terminal conservada, una región no conservada que contiene la señal de localización nuclear (NLS) y un dominio C-terminal muy conservado del tipo *Plant HomeoDomain* (PHD). En la figura 2 se muestran las secuencias de aminoácidos de las cinco proteínas ING humanas y las regiones que se describen a continuación.

En el extremo N-terminal se encuentra un dominio LZL del tipo cremallera de leucinas (*LeucineZipperLike*) presente en todas las proteínas ING, excepto en ING1 (en ninguna de sus diferentes variantes). Este dominio, con residuos de leucina/isoleucina conservados podría funcionar como sitio de homodimerización o heterodimerización de las ING con otras proteínas que también contengan este tipo de dominios, como algunos factores de transcripción. En el caso de ING2 esta región es necesaria para su interacción con p53 y la inducción de la apoptosis (22). En la región N-terminal, ING1 posee una secuencia de interacción con la proteína PCNA (antígeno nuclear de células en proliferación, es un cofactor de la ADN polimerasa δ que favorece la síntesis de ADN, pudiendo también actuar en

[1] Las variantes más abundantes de ING1, ING3 e ING4 son p33ING1b, p47ING3 y p29ING4v1, respectivamente. A ellas se refieren los nombres ING1, ING3 e ING4 de aquí en adelante.

reparación de ADN cuando existe un daño en el ADN), a través de la cual promueve la participación de PCNA en reparación de ADN, en respuesta a un daño en el ADN inducido por radiación ultravioleta (UV) (7).

A continuación, todos los miembros de la familia ING poseen una región muy conservada llamada dominio potencial de regulación de cromatina (PCR) caracterizada por el motivo de secuencia KIQI (ING1/2) o KVQL (ING3/4/5) (5). Otros residuos conservados de este dominio hacen que ING3 diverja de ING4-5. Esta región parece estar implicada en la unión a complejos de acetilación (HAT) y desacetilación (HDAC) de histonas durante los procesos de remodelación de la cromatina y regulación de la expresión génica (9). Así, se ha descrito mediante ensayos de inmunoprecipitación de extractos celulares y de *pull-down* con las proteínas aisladas, que ING1 interacciona con la proteína SAP30 como parte del complejo mSin3/HDAC1/2 de desacetilación de histonas a través de su región N-terminal (englobando los dominios LZL y PCR) (9). Esta participación de ING1 en la desacetilación de histonas se ha propuesto como explicación del papel de ING1 en la formación de heterocromatina que acompaña a la inducción de senescencia celular (14). ING2 interacciona con otro complejo HDAC mientras que

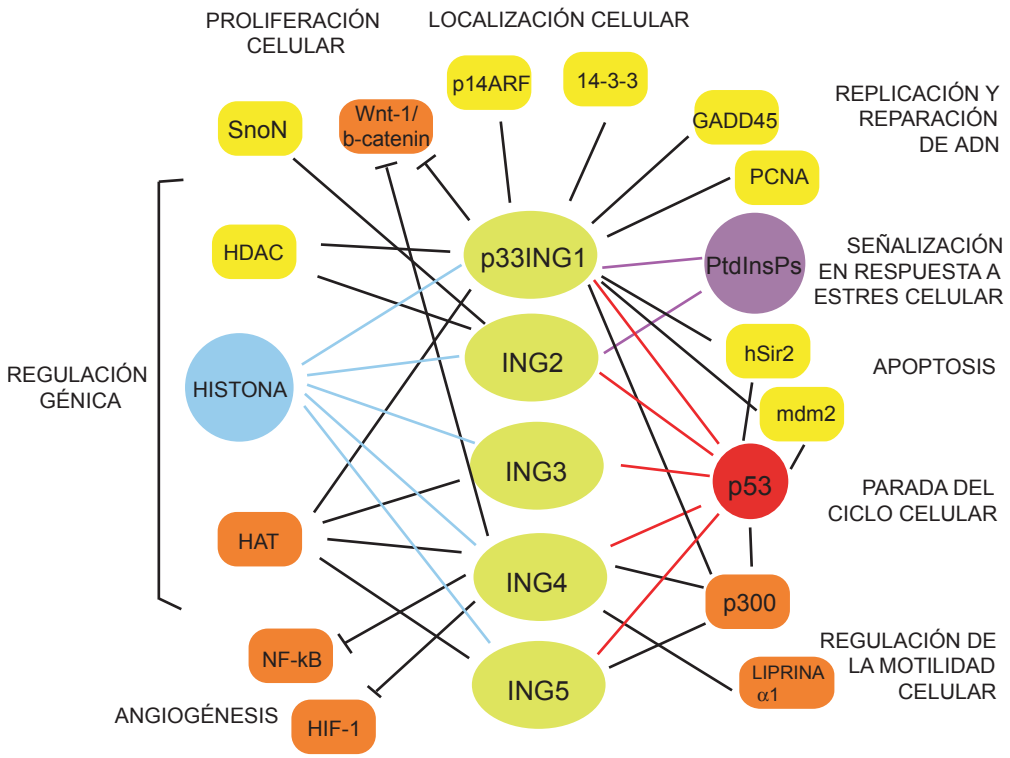

Fig. 1. Resumen (no exhaustivo) de la amplia red de proteínas y lípidos que interaccionan con los miembros de la familia ING (representados con óvalos en verde) y los diversos procesos celulares en los que están implicados estas interacciones. Las moléculas representadas con círculos indican las proteínas o lípidos relacionados con ING4 con las que se ha trabajado a lo largo de esta tesis (p53, en rojo; PtdInsPs, en morado; histonas, en azul), las coloreadas en amarillo representan proteínas que interaccionan con miembros de la familia ING, pero no con ING4, y en naranja las que aun interaccionando con ING4 no han sido estudiadas en esta tesis.

ING3, ING4 e ING5 lo hacen con complejos HAT (23), aunque en estos casos se desconoce cual es la región de interacción. Curiosamente, los tres grupos en que se pueden dividir las proteínas ING atendiendo a los diferentes motivos que caracterizan al dominio PCR, al análisis filogenético y a su diferente participación en complejos HDAC (ING1 e ING2) o HAT (ING3, ING4 e ING5), son los mismos, lo cual apoya la idea de que el dominio PCR debe tener un papel importante en la especificidad de la interacción de las cinco proteínas ING con complejos de HDAC y HAT.

La región central de las proteínas ING es muy rica en residuos básicos y constituye la secuencia de localización nuclear (NLS), que contiene 3 señales de direccionamiento hacia el nucleolo (NTS) importantes para la reparación de ADN en respuesta al daño en el ADN (18). Estas tres secuencias NTS están bien definidas en ING1 e ING2, y menos definidas o ausentes en el resto de INGs. Este hecho podría indicar que ING3, ING4 e ING5 no actúan tan eficientemente en respuesta a daño en el ADN (5). En el caso de ING4, la secuencia NLS está implicada en la unión a liprina α1, regulando la propagación y la migración celular, (21) y también se ha descrito como la región de interacción directa entre p53 e ING4 (24).

Tras la NLS, en ING1 e ING2 aparece un motivo de interacción dependiente de fosforilación (PDIM) similar a los motivos canónicos de unión a las proteínas 14-3-3 RSXSphP (donde Sph indica un residuo de serina fosforilado) y RXXXSphXP (5). De hecho, se ha descrito que ING1 interacciona con las proteínas 14-3-3 de una forma dependiente de su fosforilación en el residuo S199 (20) y que esta interacción es necesaria para que ING1 aumente la expresión de $p21^{WAF1}$ en respuesta a estrés inducido por daño en el ADN. También se ha descrito que la fosforilación de ING1 en S126 (fuera del motivo PDIM) es necesaria para que ING1 ejerza su función en la supresión de la proliferación celular (25).

Cerca del extremo C-terminal se encuentra la región más conservada de las proteínas ING, el dominio PHD (26). Este dominio se ha encontrado en la mayoría de los genomas eucariotas, predominantemente en proteínas asociadas a la cromatina, y contiene un motivo C4HC3 de dedos de zinc Recientemente se ha descrito la interacción directa entre el dominio PHD de las proteínas ING y el extremo N-terminal de la histona 3 trimetilada en el residuo de K4 (27). Esto implica a la familia ING en el reconocimiento de modificaciones postraduccionales de histonas, actuando como lectores del "código de histonas" (ver más abajo, apartado 3.2.3). Además, se ha descrito al dominio PHD como necesario para el reconocimiento de la enzima HPH-2 (HIF prolil hidroxilasa responsable de la degradación por el proteosoma de HIF-1α en condiciones de normoxia a través de su hidroxilación dependiente de oxígeno) por parte de ING4. A través de la unión con HPH-2, ING4 reprime HIF-1α (factor de transcripción que induce la expresión de factores proangiogénicos) bajo condiciones de hipoxia (28, 29).

Tras el dominio PHD, en el extremo C-terminal de las INGs hay un motivo conservado llamado PBR (región polibásica), necesario para el reconocimiento de fosfatidilinositol fosfatos. Este motivo sólo está presente como tal en ING1 e ING2. Las correspondientes regiones en el resto de los miembros de la familia son también ricas en residuos básicos pero son considerablemente más cortas, y en el caso de ING4 e ING5 (contienen el motivo de secuencia RKKK) podrían funcionar como señales adicionales de localización nuclear (8). Los PtdInsPs son moléculas de señalización que regulan la supervivencia celular, el crecimiento y la proliferación (30). Aunque la mayoría de los PtdInsPs actúan en el citoplasma, en los últimos años se ha descrito su participación en rutas nucleares de transducción de señales (31, 32). De hecho, se ha descrito que ING1 e ING2 actúan como receptores nucleares de PtdIns-3-P y PtdIns-5-P (33). Esta interacción (al menos en el caso de ING2) regula la capacidad de

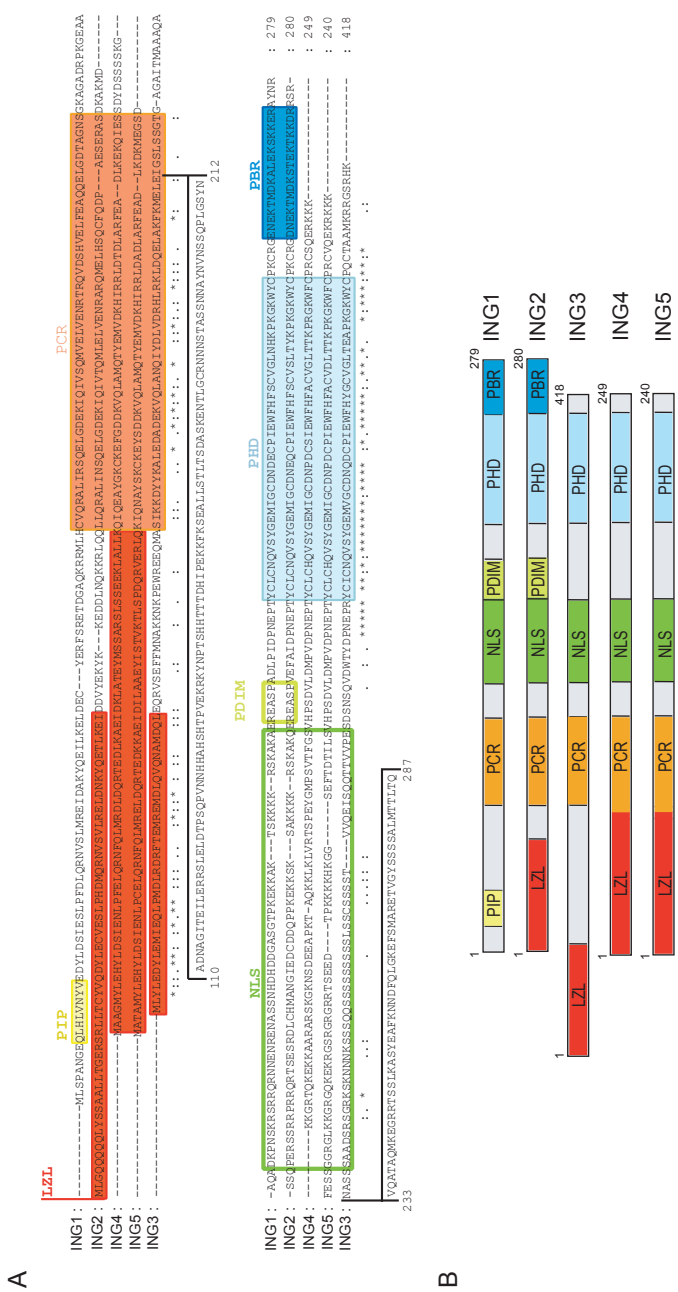

Fig. 2. Representación de los motivos de secuencia y regiones conservadas y no conservadas de las proteínas ING humanas. *A*, alineamiento de las secuencias de aminoácidos. Los diferentes dominios conservados están indicados con cuadros de colores marcando el dominio de interacción con PCNA en ING1 (amarillo), el dominio tipo cremallera de leucinas LZL (rojo), el dominio de unión a cromatina PCR (maranja), el dominio PHD (azul claro) y la región polibásica PBR (azul). La secuencia de localización nuclear (verde) y el motivo de interacción dependiente de fosforilación PDIM (verde claro) se indican con cajas vacías porque estos motivos se encuentran en una región no conservada. ING3 posee secuencias adicionales únicas que sobresalen del alineamiento (líneas negras). La conservación de la secuencia se representa mediante símbolos bajo las secuencias: los residuos de la columna son idénticos en todas las secuencias del alineamiento (*); los residuos de la columna que difieren del resto de secuencias tienen las mismas propiedades fisico-químicas (polaridad, tamaño, hidrofobicidad, aromaticidad, entre otras) (:); los residuos de la columna que difieren del resto tienen propiedades fisico-químicas cercanas, pero no idénticas (.) *B*, representación esquemática de los motivos de secuencia más importantes de las proteínas ING humanas.

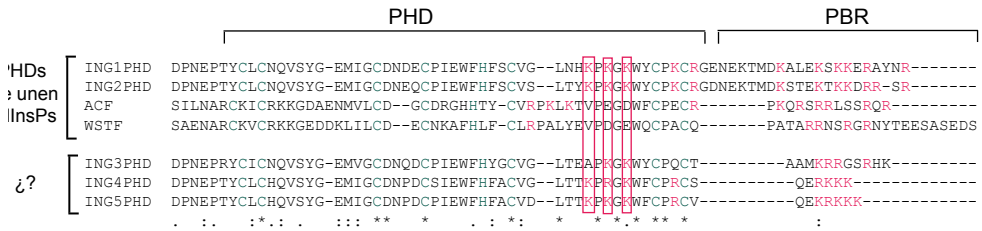

Fig. 3. Alineamiento de secuencia de los dominios PHD y motivos PBR de las proteínas ING y otras proteínas que reconocen PtdInsPs. Las secuencias se encuentran agrupadas atendiendo a su conocida (panel superior) o desconocida (panel inferior) interacción con PtdInsPs.

ING2 para unir p53 y promover apoptosis, sugiriéndose que la interacción ING2-PtdInsP podría regular los procesos biológicos que tienen lugar en el núcleo en respuesta a una situación de estrés celular (33-35). La región de interacción con PtdInsPs comprende el dominio PHD y el PBR de ING2 y, en concreto, los residuos básicos de estas regiones. Inicialmente se describió que tanto las lisinas carboxilo terminales del PHD como el PBR eran necesarios pero no suficientes para la interacción (33), pero un trabajo posterior muestra que la única región necesaria para la interacción es la PBR (36). Los dominios PHD de otras proteínas (WSTF y ACF, entre otras) reconocen PtdInsPs aunque con diferentes afinidades y especificidades, siendo esta afinidad más fuerte cuanto mayor es el número de residuos básicos del extremo C-terminal (figura 3) (33). Este hecho hace plantearse si los dominios PHD de ING3, ING4 e ING5 serán también capaces de interaccionar con PtdInsPs a pesar de las diferencias existentes en sus secuencias PBR con respecto a ING1 e ING2 (33).

2. ING4

El gen ING4 se localiza en la región cromosómica 12p13.31 y esta formado por 8 exones que codifican para una proteína mayoritariamente nuclear y minoritariamente citoplasmática de 29 kDa (5, 37). ING4 fue identificado en una búsqueda de nuevos genes supresores de tumores (38) y aislado mediante una búsqueda de genes capaces de suprimir la pérdida de inhibición por contacto sin afectar a la proliferación celular de células no cancerígenas (39, 40). ING4 está implicada, directa o indirectamente, en la remodelación de la cromatina, regulación de la propagación y migración celular, inhibición de la proliferación celular y la angiogénesis, parada del ciclo celular e inducción de apoptosis.

El gen de ING4 codifica tres variantes postranscripcionales: ING4v1 (de aquí en adelante ING4), ING4v2 e ING4v4. Las variantes ING4v2 e ING4v4 han perdido la secuencia de localización nuclear (NLS) por lo que se localizan en el citoplasma, tienen atenuada su función en inhibición de la proliferación celular y han perdido su papel en propagación celular. ING4v2 mantiene su papel en la inhibición de la migración celular, mientras que ING4v4 también lo pierde. Además, la variante ING4v4 muestra un efecto dominante negativo sobre ING4 en su capacidad para activar la transcripción del gen $p21^{WAF1}$ (gen que codifica para la proteína $p21^{WAF1}$ implicada en la regulación del ciclo celular a nivel de la fase G1) y en su función en la propagación y migración celular. Este efecto dominante no se debe

a una interacción directa entre ING4 e ING4v4, por lo que se ha propuesto que ING4v4 podría actuar secuestrando en el citoplasma proteínas con las que ING4 interacciona en el núcleo y a través de las cuales intervendría en la parada del ciclo celular (41).

La expresión de ING4 se encuentra significativamente reducida en células tumorales en comparación con células normales. Además, la falta del locus ING4 en la región cromosómica 12p13.31 se ha observado en un 10-20 % de las células de cáncer de mama, así como una pérdida de heterocigocidad acompañada de pérdida de la expresión de ING4 en células de carcinoma escamoso de cabeza y cuello (37, 39). La expresión de ING4 se encuentra reducida en gliomas y está correlacionada con el grado del tumor (12). En líneas celulares derivadas de neuroblastoma, cáncer de mama, colon y pulmón se han descrito varias mutaciones y eliminaciones parciales de ING4 (39), siendo estas modificaciones más frecuentes en la región de la NLS y en el dominio PHD (39).

ING4 inhibe la proliferación celular a través de la inhibición de la cascada de señalización celular Wnt-1/β-catenin (42). También se ha observado que ING4 aumenta la sensibilidad a agentes que dañan el DNA, lo que implica que ING4 podría ser un buen candidato para terapia génica en combinación con quimioterapia (42, 43). A diferencia de otras INGs, ING4 no parece actuar en la regulación de la reparación de ADN. La sobreexpresión de ING4 provoca: la inducción de apoptosis (43), el descenso del número de células RKO en fase S (38), la activación de la transcripción de $p21^{WAF1}$ (38) y la disminución en la eficiencia de formación de colonias (38), y en todos los casos de forma dependiente de p53, lo que indica que debe existir una cooperación entre ambos supresores tumorales.

Ensayos de inmunoprecipitación con extractos celulares y de *pull-down* con las proteínas aisladas indican que ING4 interacciona con p53, siendo esencial para la unión la secuencia NLS de ING4 (24). ING4 también interacciona con p300 (proteína que participa en acetilación de factores de transcripción como p53 y en co-activación de la transcripción a través de su actividad HAT), de modo que ING4 haría de intermediario entre p300 y p53, induciendo la acetilación de p53 en el residuo K382 y consiguiente estabilización y activación de p53 (38).

2.1 ING4 en angiogénesis, crecimiento tumoral e invasión.

La angiogénesis es la formación de nuevos vasos sanguíneos a partir de otros preexistentes. En el crecimiento tumoral, las células cancerosas segregan factores proangiogénicos que inducen la vascularización necesaria para el crecimiento del tumor.

En glioblastoma, se han descrito niveles de expresión muy reducida de ING4 que correlacionan con un aumento importante del volumen vascular y del crecimiento del tumor. La inhibición de la angiogénesis por parte de ING4 se produce a través de su interacción con la subunidad p65 del factor de transcripción NF-κB, un factor de transcripción que activa genes proangiogénicos. La interacción inhibe NF-κB y con ello, la vascularización del tumor y su crecimiento (12, 44, 45).

En condiciones de hipoxia (bajos niveles de oxígeno), se ha observado que ING4 modula la actividad de HIF-1α (factor de transcripción inducible por hipoxia). HIF-1α induce la expresión de factores proangiogénicos y promueve la progresión del tumor a través de vías de adaptación a hipoxia (angiogénesis, transporte de oxígeno y modificación del metabolismo de la glucosa, entre otras) (46, 47). Bajo condiciones de hipoxia, la asociación de HIF-1α con HPH-2 permite el reclutamiento de ING4

(a través de su dominio PHD), que a su vez inhibe la actividad transcripcional de HIF-1α (28, 29, 48), colaborando en la inhibición de la progresión del tumor.

ING4 suprime ciertos rasgos típicos de las células tumorales, tales como la pérdida de inhibición por contacto y el crecimiento celular en *soft agar*, pero no afecta directamente a la regulación del ciclo celular, a pesar de provocar indirectamente una disminución del número de células en fase S. Además, ING4 afecta a la proliferación celular en líneas celulares de cáncer pero no en células no tumorales lo que podría ser una oportunidad para el desarrollo de nuevas terapias de cáncer sin toxicidad para las células no tumorales. Una posible explicación de cómo ING4 puede actuar en la supresión tumoral sin afectar directamente al ciclo celular, es que regule la transcripción de genes implicados en el control de múltiples funciones celulares (39, 40).

Por otro lado, la célula tumoral posee la capacidad de propagarse y migrar y, con ello, provocar metástasis. ING4 actúa en el citoplasma regulando la propagación y migración celular (21, 41), aunque se desconoce su efecto concreto sobre las propiedades de invasión y metástasis de células tumorales.

3. Regulación de la estructura de la cromatina y el código de histonas.

En los organismos eucariotas, el metabolismo del ADN (replicación, reparación y transcripción) se ve afectado por el grado de empaquetamiento del mismo en la cromatina. La unidad básica de la cromatina es el nucleosoma, en el que dos copias de cada una de las cuatro histonas centrales (H2A, H2B, H3 y H4) forman un octámero alrededor del cual se enrollan firmemente 146 pares de bases de ADN (figura 4). Las histonas son pequeñas proteínas básicas conservadas a lo largo de la evolución que poseen una región C-terminal globular imprescindible para la formación del nucleosoma y un extremo N-terminal cargado y flexible (49). La cromatina puede adoptar dos formas: la heterocromatina, que se caracteriza por su alto grado de condensación, implica una baja accesibilidad de los genes y una baja o nula transcripción de los genes en los segmentos de ADN empaquetados de esta forma, y la eucromatina que posee un menor grado de condensación, por lo que sus genes están más accesibles, relacionándose con replicación y transcripción activas. Los procesos de remodelación de la cromatina actúan modificando la estructura de la misma, provocando así una mayor o menor accesibilidad al ADN.

Un mecanismo molecular clave en la remodelación de la cromatina es el que tiene lugar a través de modificaciones postraduccionales de las colas N-terminales de las histonas el nucleosoma. Estas modificaciones de histonas, junto con la metilación del ADN constituyen los cambios fenotípicos heredables en una célula u organismo que no se deben a cambios en la secuencia de nucleòtidos del ADN, y a lo que se denomina herencia epigenética.

Las modificaciones postraduccionales descritas en las histonas son: metilación, acetilación, fosforilación, ubiquitinización y sumoilación. Estas modificaciones regulan la accesibilidad de los genes, afectando directamente a la estructura de la cromatina o indirectamente, a través del reclutamiento de complejos de remodelación de la cromatina (figura 5). Así, la acetilación en los residuos de lisina de las colas de histonas neutraliza su carga positiva, reduciendo las interacciones electrostáticas entre el ADN y las histonas. Con esto, el equilibrio entre histona libre e histona empaquetada en nucleosomas se desplaza a favor de las histonas libres, de modo que la cromatina adopta, localmente, una estructura menos condensada, facilitando de esta forma la activación de la transcripción (3, 50). Las modificaciones de

las histonas son reversibles e interdependentes, y el conjunto es dinámico y cambiante. La presencia de ciertas modificaciones en residuos concretos facilita o bloquea la modificación de otros (51). Así por ejemplo, la fosforilación de la histona 3 en el residuo de S10 facilita la metilación de K4, y la acetilación de K9 y de K14. La acetilación de K9 dificulta su metilación, característica de formación de heterocromatina (52), resultando en una conformación más abierta de la cromatina y promoviendo de esta forma la activación de la transcripción

Al conjunto de modificaciones postraduccionales que se producen en las colas de histonas y cómo estas modificaciones intervienen en la regulación de la reparación, replicación y transcripción génica se le conoce como hipótesis del "código de histonas"(49, 53).

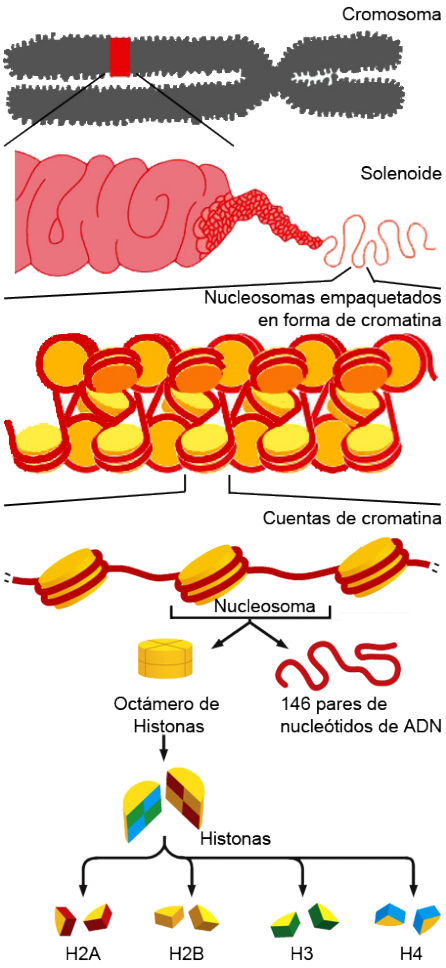

Fig. 4. Representación de la unidad mínima de la cromatina y sus grados de empaquetamiento. Este esquema muestra los diferentes grados de empaquetamiento de la cromatina desde el mayor grado de condensación representado por el cromosoma hasta la unidad mínima de la cromatina, el nucleosoma (compuesto del octámero de histonas y 146 pares de bases de ADN). Figura adaptada a partir de Molecular Biology of the Cell, Bruce Alberts et al. 5th ed. (2008) p212, 244

El código de histonas lo "escriben" complejos enzimáticos específicos y lo "leen" dominios de reconocimiento también específicos de cada tipo de modificación. La metilación y acetilación la llevan a cabo complejos de metiltransferasas de histonas (HMT) y demetilasas de histonas (HDM), y complejos de acetiltransferasas (HAT) y deacetilasas de histonas (HDAC), respectivamente. También existen enzimas encargadas de la fosforilación, ubiquitinización y sumoilación de histonas (50, 54). Mientras que los HMT y HDM tienen implicaciones en la transcripción génica muy diversa en función del residuo modificado, los HAT y HDAC ("estritores") actúan, por lo general, como co-activadores y co-represores de la transcripción génica en combinación con factores de transcripción que les dotan de un mayor grado de especificidad. Las moléculas "lectoras" se encargan del reconocimiento específico de las modificaciones de histonas, reclutando a los complejos "escritores" (HAT, HDAC y HMT, entre otras) hacia determinadas posiciones de la cromatina e iniciando respuestas biológicas como la activación/represión de la transcripción.

Existen evidencias contrastadas que implican a las proteínas ING en la regulación de la estructura de la cromatina y la lectura del código de histonas a través de su asociación física con los complejos

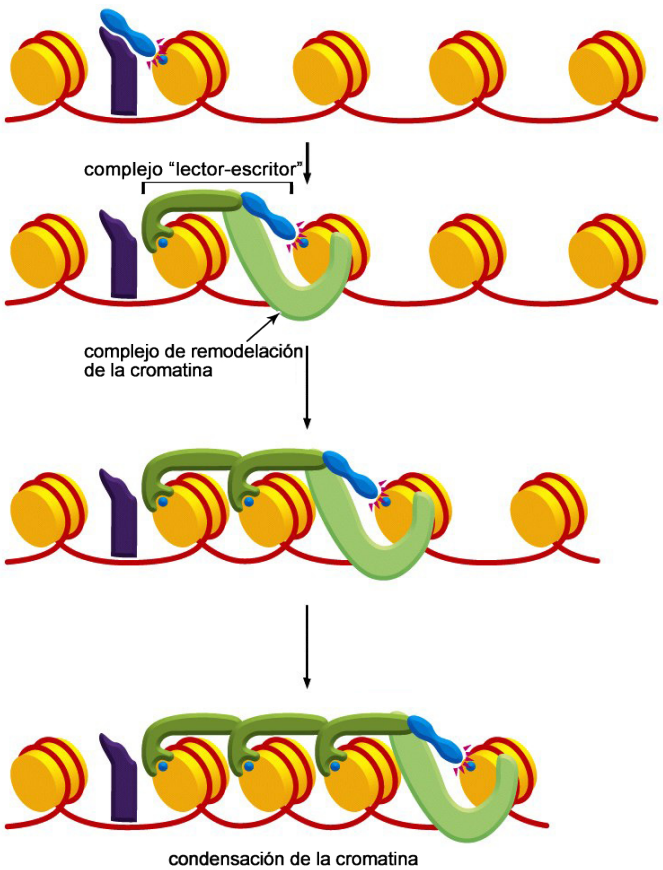

Fig 5. Modelo de propagación de cambios en la estructura de la cromatina. Las moléculas lectoras, escritoras y proteínas remodeladoras de la estructura de la cromatina se unen a modificaciones específicas de las histonas (en este caso marcas que implican represión transcripcional) propagando la condensación de la cromatina. Figura adaptada a partir de Molecular Biology of the Cell, Bruce Alberts et al. 5th ed. (2008) p228

enzimáticos que regulan la acetilación de histonas y con las colas modificadas de histonas, llegándose a proponer a estas asociaciones como la causa última por la cual las proteínas ING están implicadas en funciones biológicas tan diversas (3, 34, 35, 55).

3.1 Complejos de acetilación y desacetilación de histonas.

La acetilación tiene lugar en residuos de lisina de las cuatro histonas y la llevan a cabo las enzimas (HAT)[2], las cuales se encuentran formando parte de grandes complejos multiproteicos que les dirigen hacia dominios cromosómicos específicos. Estos enzimas pueden clasificarse en familias homólogas con diferentes mecanismos de catálisis y de unión a sustrato (50). Estructuralmente, las proteínas contienen un dominio con actividad HAT, seguido de uno o más bromodominios, dominios ricos en cisteínas e histidinas (del tipo TAZ, PHD y ZZ) y cromodominios (50). Estos dominios están involucrados en el reconocimiento de histonas modificadas (bromodominios, cromodominios y PHD) e interacciones con otras proteínas como factores de transcripción (dominios ricos en cisteínas e histidinas). Debido a esto, se especula que los complejos HAT puedan ser reclutados a la cromatina gracias al reconocimiento de histonas metiladas y/o acetiladas por parte de cromodominios, bromodominios o PHD. Los complejos HAT están relacionadas con activación de la transcripción con la excepción de HBO1 (acetiltransferasa de histonas de unión a ORC-1), el cual está implicado en represión y activación de la transcripción.

La acetilación de lisinas es un proceso reversible gobernado por la acción opuesta de las actividades HAT y HDAC. Como las HATs, las HDACs se encuentran formando parte de grandes complejos multiproteicos pero, a diferencia de las HATs, están generalmente involucradas en la represión transcripcional.

Se ha descrito mediante ensayos de *TAP-tag* (ING1-5) e inmunoprecipitación en células vivas (en el caso de ING1) que las proteínas de la familia ING se asocian físicamente con complejos HAT y HDAC

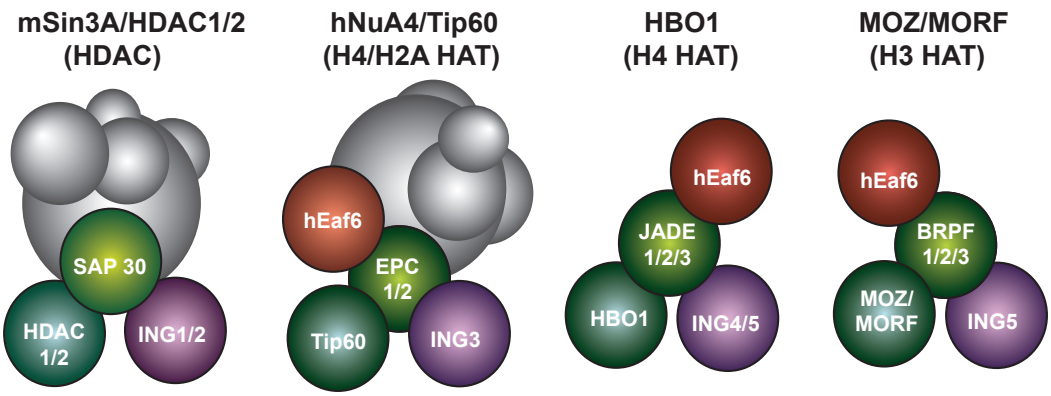

Fig. 6. Representación esquemática de los diferentes complejos HAT y HDAC de los que forman parte las proteínas ING humanas. Se representan en gris los componentes desconocidos o variables de los complejos, en morado las diferentes INGs y en otros colores los diferentes componentes de identidad conocida.

[2]Las siglas HAT y HDAC se usan para denominar a los complejos, proteínas o dominios con actividad enzimática de acetilación o desacetilación de histonas, respectivamente

(figura 6). Las INGs se podrían clasificar desde el punto de vista de su participación en estos complejos de remodelación de la cromatina, en tres grupos: ING1 e ING2, ING3, e ING4 e ING5. Las proteínas ING1 e ING2 forman parte del complejo mSin3/HDAC1/2 participando de esta manera en represión transcripcional. En el caso de ING1, es el dominio N-terminal (LZL y PCR) y en concreto sus 99 aminoácidos iniciales, los responsables de la interacción (9, 23, 56). ING3 se encuentra asociado con el complejo HAT llamado hNuA4/Tip60 (que acetila las histonas 2A y 4) (23) e ING5 forma parte de dos tipos de complejos HAT: HBO1 (que acetila la histona 4) y MOZ/MORF (acetila la histona 3) (23). ING4 comparte con ING5 la participación en complejos de acetilación HBO1, pero no en MOZ/MORF (23). La región de ING4 propuesta para interaccionar con el complejo HBO1 en base a los estudios realizados con ING1, son los dominios LZL y/o PCR de la región N-terminal (5, 55).

La proteína HBO1 es la subunidad catalítica dentro del complejo que lleva su nombre y se encarga de la acetilación específica de la histona 4 en K5, K8 y K12. La proteína HBO1 es la única proteína con actividad HAT involucrada en represión transcripcional (57-59). Recientemente, se le ha involucrado también en la activación de la transcripción (60). La región responsable de la función represora de la proteína HBO1, es su dominio N-terminal rico en serinas, que no comparte con el resto de proteínas HAT (59, 61). En relación a esto, se ha descrito que HBO1 puede actuar como co-represor del factor de transcripción NF-κB y otros factores nucleares (58, 59) e interaccionar con p53, provocando la disminución de su actividad HAT (62).

No esta claro si existe una interrelación entre la inhibición de NF-κB por parte de ING4 y el papel de HBO1 en represión de NF-κB, ni tampoco si la interacción de ING4 con p53 tiene alguna relación con la disminución de actividad HAT de HBO1 cuando este interacciona con p53, pero se ha propuesto que ING4 podría tener una función "puente" reclutando el complejo HBO1 hacia sitios transcripcionalmente activos marcados por la presencia de factores de transcripción específicos como NF-κB y p53 (34, 35, 55).

3.2 "Lectura" del código de histonas

Las diferentes modificaciones postraduccionales que sufren las colas de las 4 histonas que forman parte de los nucleosomas se muestran esquemáticamente en la figura 7A.

La acetilación de histonas tiene lugar en los grupos amino de residuos de lisina de las cuatro histonas y está relacionada con activación de la transcripción (50). La fosforilación, ocurre principalmente en la histona 3, pudiendo tener efectos tanto en activación como en represión de la transcripción. Los grupos amino de residuos de lisina también pueden ser ubiquitinilados o sumoilados, pero debido al gran tamaño de estas modificaciones no se sabe si afectan directamente a la estructura de la cromatina, si promueven o inhiben la interacción con otras proteínas, o ambas cosas (63). La metilación de histonas se produce en residuos de lisina y arginina, siendo las histonas 3 y 4 las más frecuentemente metiladas. La arginina puede ser mono- o dimetilada simétrica o asimétricamente en su grupo guanidinio. La lisina puede aceptar uno, dos o tres grupos metilos en su grupo amino, generando las formas mono-, di-, o trimetilada. Se ha encontrado que su distribución a lo largo de los genes no es homogénea, siendo más abundante H3K4me3 en el extremo 5´, H3K4me2 en la región intermedia del gen y H3K4me en el extremo 3´(64). Existe una correlación entre metilación y sitios activos en transcripción, aunque su reconocimiento puede tener como consecuencia tanto la activación como la represión de la transcripción (50, 65-68).

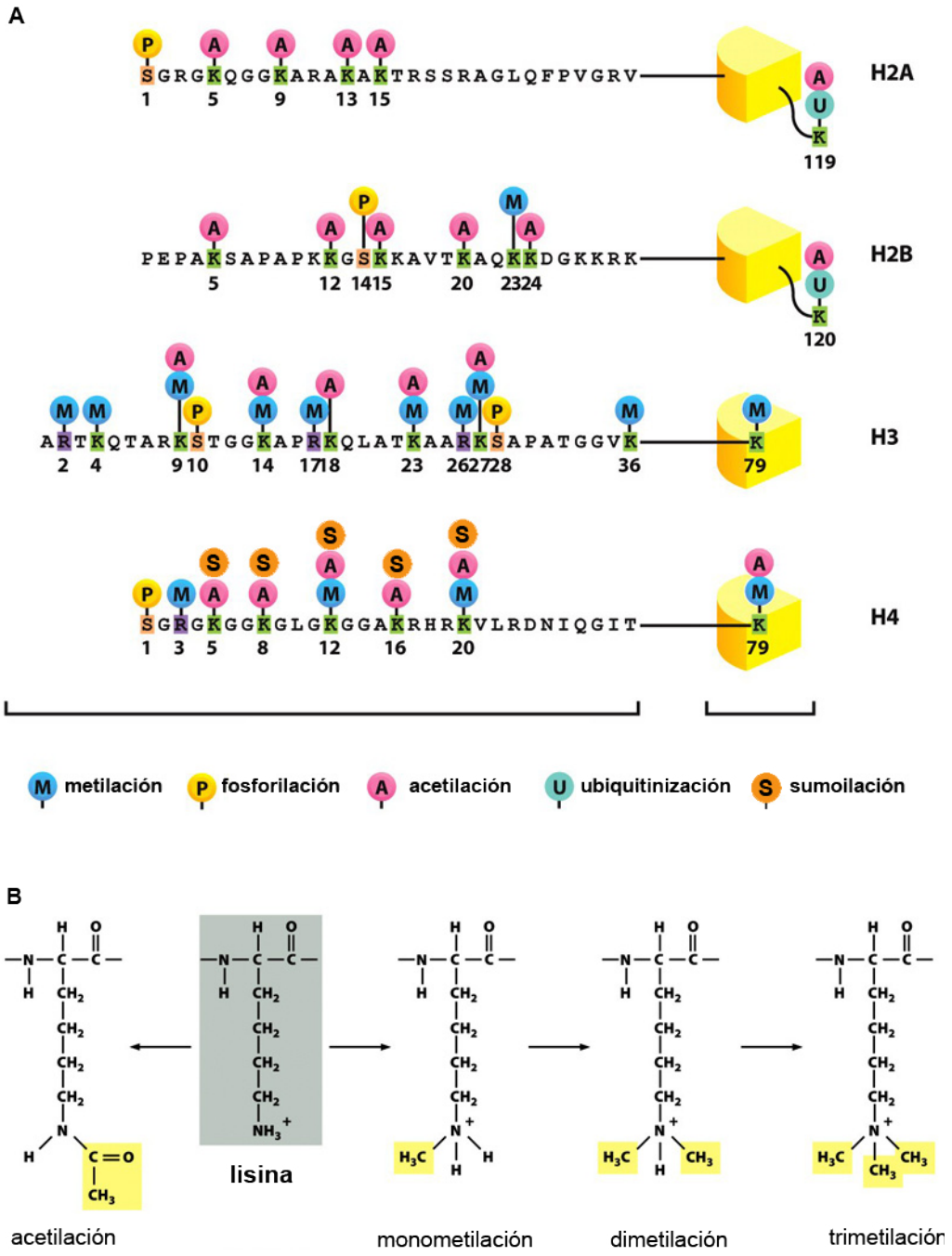

Fig. 7. Modificaciones postraduccionales de las colas N-terminales de histonas. *A,* Modificaciones postraduccionales descritas para las cuatro histonas que componen los nucleosomas. Las diferentes modificaciones, así como los diferentes residuos que las sufren, se representan con distintos colores especificados en el diagrama. *B*, Acetilación y diferentes estados de metilación de lisinas en las colas de histonas. Figura adaptada a partir de Molecular Biology of the Cell, Bruce Alberts et al. 5th ed. (2008) p222-223

Los lectores del código de histonas son dominios que reconocen específicamente las modificaciones postraduccionales de las histonas. Este reconocimiento lo realizan de un modo dependiente tanto del residuo y la modificación concreta, como de la secuencia particular de la histona modificada. Se conocen varios tipos de dominios lectores, siendo los mejor caracterizados los bromodominios, cromodominios y dominios PHD (69).

3.2.1 Reconocimiento de acetil-lilsinas: Bromodominios

Los bromodominios son pequeñas dominios que funcionan como módulos de reconocimiento de histonas acetiladas. La estructura de los bromodominios está formada por 4 hélices empaquetadas con dos bucles de distinta longitud entre sus hélices. Este empaquetamiento genera un bolsillo hidrofóbico estrecho y profundo sobre el conjunto de hélices, en el que se acopla la larga cadena lateral de acetil-lisina (como en el caso del bromodominio de de la proteína Gcn5p representada en la figura 8A) (70). En proteínas que constan de dos bromodominios (como el caso de TAF-1), estos dos módulos actúan en tandem generando un sitio de reconocimiento de histonas diacetiladas (H4K5acK12ac) en la interfaz de los dos dominios (70).

3.2.2 Reconocimiento de metil-lisinas

Los diferentes estados de metilación confieren diferentes propiedades fisicoquímicas a las lisinas, lo que hace posible que sean diferenciados por los dominios lectores. Todos los estados de metilación de la cadena lateral de la lisina están cargados positivamente a pH fisiológico, pero con la adición de grupos metilo, la hidrofobicidad y el radio del catión del grupo metilamonio aumenta y, con ello, la capacidad para formar enlaces de hidrógeno decrece (figura 7B).

Los estudios recientes acerca del reconocimiento de metil-lisinas por parte de dominios lectores sugieren la existencia de mecanismos generales mediante los que estos dominios se unen a histonas con especificidad del estado de metilación y de secuencia. Estos estudios se han realizado con péptidos que mimetizan la cola flexible de la histona. El reconocimiento de los estados de metilación de la lisina se basa en la formación de interacciones específicas entre la lisina metilada y residuos aromáticos de la proteína que forman una "jaula" alrededor del metilamonio de la lisina, según se observa en las estructuras de los complejos determinados hasta el momento, algunos de los cuales se muestran en la figura 8.

En el caso de los dominios que reconocen lisina trimetilada, la jaula aromática forma una cavidad estrecha muy sensible y específica al tamaño de la modificación. Entre los residuos aromáticos y el amonio cuaternario se forman interacciones de tipo catión-π. Estas interacciones son de naturaleza predominantemente electrostática, produciéndose entre un catión y un campo con potencial electrostático negativo generado por el momento cuadrupolar del anillo aromático (71).

A partir de las estructuras conocidas, y en especial de las estructuras de los dominios libres y unidos a los péptidos, se ha concluido que, mientras que las colas flexibles de las histonas pueden adquirir una estructura definida al formar el complejo, los dominios actúan como armazones estáticos que, salvo alguna excepción, apenas sufren modificaciones con la unión de la histona. Este hecho minimizaría el coste de entropía conformacional de la unión, favoreciendo la formación del complejo (69).

La discriminación entre los distintos estados de metilación viene dada por la forma del sitio de unión en el dominio. Mientras que el reconocimiento de la forma trimetilada se produce mediante una cavidad, cuando los estados de metilación son menores los sitios de unión se hacen más superficiales, anchos y accesibles, adquiriendo la forma de un surco en la superficie del lector sobre el que yace la lisina modificada. Las diferentes secuencias de las histonas son discriminadas por los dominios lectores mediante dos aspectos de la unión: la complementariedad de forma e interacciones que tienen lugar cuando la histona se une a la proteína, y la longitud del segmento N-terminal hasta la lisina metilada. La cola de la histona adquiriere una conformación extendida de hebra-β que se alinea de forma antiparalela con las hebras-β de la proteína. Este alineamiento, estabilizado por una red de enlaces de hidrógeno e interacciones electrostáticas, provoca que la metil-lisina y los residuos en posiciones +2 y -2 con respecto a ella adopten una forma complementaria con los surcos en la superficie del dominio. La participación del extremo N-terminal de la histona en la interacción es importante en los casos en los que está cerca

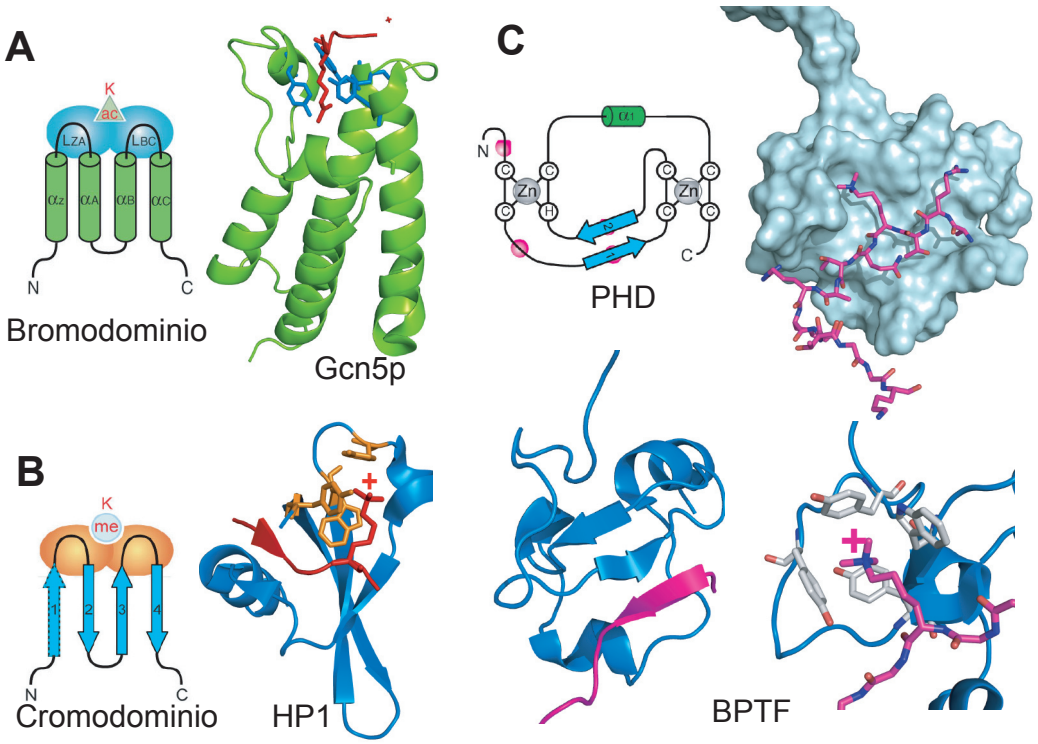

Fig. 8. Representación de los módulos lectores del código de histonas mejor caracterizados (Bromodominio, cromodominio y PHD) y estructura de los complejos que forman con histonas modificadas. *A*, esquema de la estructura de bromodominios y ejemplo del reconocimiento de lisinas acetiladas: complejo Gcn5p-H4K16ac (código del PDB:1e6i). *B*, esquema estructural de los cromodominios y ejemplo del reconocimiento de lisinas metiladas: complejo HP1-H3K9me3 (código del PDB: 1kne). *C*, esquema de la estructura de PHD, ejemplo del reconocimiento de lisinas metiladas: complejo BPTF-H3K4me3 (código del PDB: 2fuu) y detalle de la unión del metilamonio de la lisina4 en la histona 3 en la jaula aromática de BPTF (de izquierda a derecha). En los esquemas, los cilindros se refieren estructura secundaria de hélice-α, mientras que las flechas son hebras-β (también en las estructuras). Las estructuras se representan con diagrama de cintas, a excepción de los residuos que forman la jaula aromática y el residuo metilado o acetilado de la histona, que se representan con varillas.

de la lisina metilada, como ocurre en el caso de la histona 3 metilada en K4. En estos casos, el extremo N-terminal cargado positivamente se ancla en una región del dominio que actúa como barrera para la extensión N-terminal característica de histonas con lisinas modificadas más alejadas en la secuencia.
Se han descrito dos clases de dominios con la propiedad de unirse a lisinas metiladas: cromodominios y dominios PHD.

3.2.2.1 Cromodominios

Los cromodominios son pequeños módulos de reconocimiento de metil-lisinas. La estructura de los cromodominios se caracteriza por tener una lámina-β antiparalela de 3 hebras acompañada de una hélice-α que cruza la lámina (figura 8B). El sitio de unión para lisinas metiladas está formado por las cadenas laterales de tres residuos aromáticos. Un ejemplo de reconocimiento de lisinas metiladas por cromodominios es el de la proteína HP1 (proteína de heterocromatina 1 implicada en la represión de la transcripción mediante la formación de heterocromatina) que reconoce la histona 3 metilada en el residuo K9. El cromodominio de HP1 se une a las formas di- y trimetilada com más afinidad que a la monometilada (K_D=2.5, 7.0 y 1,900 μM para las formas tri-, di-, y monometiladas, respectivamente). En el complejo se forma un barril β de cinco hebras (cuatro hebras de HP1 más una del péptido) representado en la figura 8B (72-74). El papel más importante en el reconocimiento de histonas por cromodominios lo tiene la interacción tipo catión-π y no el efecto hidrofóbico (75). Otro ejemplo de reconocimiento es el del doble cromodominio de la proteína CHD1 (proteína cromo-ATPasa-helicasa de unión a ADN, regula la transcripción génica a través de la modificación de la estructura de la cromatina) que reconoce histona 3 metilada en K4, discriminando poco entre los tres estados de metilación (K_D = 5.0, 5.0 y 17.0 μM para las formas tri-, di- y monometilada, respectivamente) (76, 77) . Además, CHD1 es capaza de reconocer la histona 3 metilada en K4 en combinación con otras modificaciones como trimetilación o acetilación en K9, o fosforilación en S10 con afinidades similares, pero muestra una drástica disminución de la afinidad cuando se combina con fosforilación en T3 (76). El hecho de si un lector puede funcionar como dominio de reconocimiento simultaneo de dos modificaciones de histonas, o si varios lectores de una misma proteína actúan de forma cooperativa en la lectura de una única o varias histonas, se perfila como un complejo e importante campo para la investigación futura.

3.2.2.2 *Plant HomeoDomains* (PHD)

Estos dominios, de aproximadamente 60 aminoácidos, tienen un motivo C_4HC_3 (4 cisteínas, 1 histidina, 3 cisteínas) con un espaciado característico entre estos residuos. Este motivo de cisteínas e histidina coordina 2 iones Zn^{2+} que mantienen la estabilidad de la estructura. Los dominios PHD poseen muy pocos elementos de estructura secundaria, salvo dos cortas hebras-β antiparalelas y una vuelta de hélice-α, y tienen varios bucles de longitud variable e importantes para la especificidad de reconocimiento de los diferentes PHD (26, 78-80) (figura 8C).
Los dominios PHD son módulos de reconocimiento de histona 3 en K4. Así, la proteína BPTF (factor de transcripción bromodominio-PHD, participa en la regulación de la transcripción) une histona 3 metilada en K4 a través de su dominio PHD, y forma parte del complejo de remodelación de la cromatina NURF (factor de remodelación del nucleosoma, necesario para la remodelación de cromatina e involucrado

en activación de la transcripción). BPTF une H3K4me3/2 (K_D=2.7 y 5.0 μM respectivamente), discriminando frente a las formas monometilada y no metilada de histona 3 en K4 mediante una jaula de cuatro residuos aromáticos (un triptófano y tres tirosinas) que forma una cavidad en la que se inserta la lisina metilada (figura 8C).

El PHD de la proteína Spp1 (subunidad del complejo de metilación de histonas Set1p) reconoce la dimetilación de histona 3 en K4 y este reconocimiento es inhibido por la dimetilación simultánea en R2. Esta modificación evita la unión del complejo Set1p y consiguiente metilación de histona 3, con el fin de mantener silenciados determinados genes de la cromatina (81).

3.2.3 Las proteínas ING en el reconocimiento de histonas

Al inicio de esta tesis, la información acerca de las proteínas ING era mucho más limitada en cuanto a funcionalidad y conocimiento estructural. De hecho, toda la información que ahora involucra a la familia ING en la lectura del código de histonas era desconocida, ignorándose incluso que los dominios PHD pudiesen reconocer histonas metiladas. Durante el transcurso de la tesis se describieron las estructuras de los dominios PHD de BPTF, ING1, ING2 e ING5 en complejos con péptidos que mimetizan las colas N-terminales flexibles de histona 3 metilada en lisina 4.

El primer miembro de la familia en ser estudiado en profundidad fue el dominio PHD de ING2 que, a pesar de estar involucrado en represión de la transcripción formando parte del HDAC mSin3/HDAC1/2, interacciona con H3K4me3 (27, 82).

El PHD de ING2 interacciona específicamente con H3K4me3 mostrando una afinidad uno y dos órdenes de magnitud mayor para H3K4me3 (K_D= 208.0 μM) que para H3K4me2 (K_D= 15.0 μM) y H3K4me1 (K_D= 1.5 μM), pero no reconoce ni la forma no metilada de H3, ni otras formas metiladas como histona 3 metilada en K9 o histona 4 metilada en K20 (82). En el reconocimiento está presente la interacción tipo catión-π entre el grupo metilamonio de la lisina y la jaula aromática que, a diferencia del PHD de BPFT, está formada por sólo dos residuos aromáticos, y las otras dos posiciones están ocupadas por una serina y una metionina (27, 82).

En los complejos formados por los dominios PHD de ING1 e ING5 con H3K4me3 se observa el mismo tipo de reconocimiento que en el caso de ING2 (82-84).

La observación de que las funciones de reparación de ADN y apoptosis requieren de la unión específica del PHD a H3K4me3 y el hecho de que ING1 está involucrada en la remodelación de la cromatina a través de complejos mSin3/HDAC1/2 son las bases sobre las que se propone que el reconocimiento de H3K4me3 por ING1 podría reclutar los correspondientes complejos HDAC de los que forman parte, que desacetilarían ciertas posiciones de las histonas, facilitando así la represión de la transcripción de genes que conlleven como efecto final la inducción de apoptosis y reparación de ADN (82, 84).

Bibliografía

1. Harris, H., Miller, O. J., Klein, G., Worst, P., and Tachibana, T. Suppression of malignancy by cell fusion. Nature, 223: 363-368, 1969.
2. Garkavtsev, I., Kazarov, A., Gudkov, A., and Riabowol, K. Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. Nat Genet, 14: 415-420., 1996.
3. Gong, W., Suzuki, K., Russell, M., and Riabowol, K. Function of the ING family of PHD proteins in cancer. Int J Biochem Cell Biol, 37: 1054-1065, 2005.
4. Ythier, D., Larrieu, D., Brambilla, C., Brambilla, E., and Pedeux, R. The new tumor suppressor genes ING: genomic structure and status in cancer. Int J Cancer, 123: 1483-1490, 2008.
5. He, G. H., Helbing, C. C., Wagner, M. J., Sensen, C. W., and Riabowol, K. Phylogenetic analysis of the ING family of PHD finger proteins. Mol Biol Evol, 22: 104-116, 2005.
6. Cheung, K. J., Jr. and Li, G. p33(ING1) enhances UVB-induced apoptosis in melanoma cells. Exp Cell Res, 279: 291-298, 2002.
7. Scott, M., Bonnefin, P., Vieyra, D., Boisvert, F. M., Young, D., Bazett-Jones, D. P., and Riabowol, K. UV-induced binding of ING1 to PCNA regulates the induction of apoptosis. J Cell Sci, 114: 3455-3462, 2001.
8. Feng, X., Hara, Y., and Riabowol, K. Different HATS of the ING1 gene family. Trends Cell Biol, 12: 532-538, 2002.
9. Kuzmichev, A., Zhang, Y., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). Mol Cell Biol, 22: 835-848, 2002.
10. Loewith, R., Meijer, M., Lees-Miller, S. P., Riabowol, K., and Young, D. Three yeast proteins related to the human candidate tumor suppressor p33(ING1) are associated with histone acetyltransferase activities. Mol Cell Biol, 20: 3807-3816, 2000.
11. Vieyra, D., Loewith, R., Scott, M., Bonnefin, P., Boisvert, F. M., Cheema, P., Pastyryeva, S., Meijer, M., Johnston, R. N., Bazett-Jones, D. P., McMahon, S., Cole, M. D., Young, D., and Riabowol, K. Human ING1 proteins differentially regulate histone acetylation. J Biol Chem, 277: 29832-29839, 2002.
12. Garkavtsev, I., Kozin, S. V., Chernova, O., Xu, L., Winkler, F., Brown, E., Barnett, G. H., and Jain, R. K. The candidate tumour suppressor protein ING4 regulates brain tumour growth and angiogenesis. Nature, 428: 328-332, 2004.
13. Garkavtsev, I. and Riabowol, K. Extension of the replicative life span of human diploid fibroblasts by inhibition of the p33ING1 candidate tumor suppressor. Mol Cell Biol, 17: 2014-2019, 1997.
14. Abad, M., Menendez, C., Fuchtbauer, A., Serrano, M., Fuchtbauer, E. M., and Palmero, I. Ing1 mediates p53 accumulation and chromatin modification in response to oncogenic stress. J Biol Chem, 282: 31060-31067, 2007.
15. Helbing, C. C., Veillette, C., Riabowol, K., Johnston, R. N., and Garkavtsev, I. A novel candidate tumor suppressor, ING1, is involved in the regulation of apoptosis. Cancer Res, 57: 1255-1258, 1997.

16. González, L., Freije, J. M., Cal, S., López-Otín, C., Serrano, M., and Palmero, I. A functional link between the tumour suppressors ARF and p33ING1. Oncogene: in press, 2006.
17. Leung, K. M., Po, L. S., Tsang, F. C., Siu, W. Y., Lau, A., Ho, H. T., and Poon, R. Y. The candidate tumor suppressor ING1b can stabilize p53 by disrupting the regulation of p53 by MDM2. Cancer Res, 62: 4890-4893, 2002.
18. Scott, M., Boisvert, F. M., Vieyra, D., Johnston, R. N., Bazett-Jones, D. P., and Riabowol, K. UV induces nucleolar translocation of ING1 through two distinct nucleolar targeting sequences. Nucleic Acids Res, 29: 2052-2058, 2001.
19. Han, X., Feng, X., Rattner, J. B., Smith, H., Bose, P., Suzuki, K., Soliman, M. A., Scott, M. S., Burke, B. E., and Riabowol, K. Tethering by lamin A stabilizes and targets the ING1 tumour suppressor. Nat Cell Biol, 10: 1333-1340, 2008.
20. Gong, W., Russell, M., Suzuki, K., and Riabowol, K. Subcellular targeting of p33ING1b by phosphorylation-dependent 14-3-3 binding regulates p21WAF1 expression. Mol Cell Biol, 26: 2947-2954, 2006.
21. Shen, J. C., Unoki, M., Ythier, D., Duperray, A., Varticovski, L., Kumamoto, K., Pedeux, R., and Harris, C. C. Inhibitor of growth 4 suppresses cell spreading and cell migration by interacting with a novel binding partner, liprin alpha1. Cancer Res, 67: 2552-2558, 2007.
22. Wang, Y., Wang, J., and Li, G. Leucine zipper-like domain is required for tumor suppressor ING2-mediated nucleotide excision repair and apoptosis. FEBS Lett, 580: 3787-3793, 2006.
23. Doyon, Y., Cayrou, C., Ullah, M., Landry, A. J., Cote, V., Selleck, W., Lane, W. S., Tan, S., Yang, X. J., and Cote, J. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell, 21: 51-64, 2006.
24. Zhang, X., Wang, K. S., Wang, Z. Q., Xu, L. S., Wang, Q. W., Chen, F., Wei, D. Z., and Han, Z. G. Nuclear localization signal of ING4 plays a key role in its binding to p53. Biochem Biophys Res Commun, 331: 1032-1038, 2005.
25. Garate, M., Campos, E. I., Bush, J. A., Xiao, H., and Li, G. Phosphorylation of the tumor suppressor p33(ING1b) at Ser-126 influences its protein stability and proliferation of melanoma cells. Faseb J, 21: 3705-3716, 2007.
26. Bienz, M. The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci, 31: 35-40, 2006.
27. Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G., Shi, Y., Cote, J., Chua, K. F., and Gozani, O. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature, 442: 96-99, 2006.
28. Ozer, A. and Bruick, R. K. Regulation of HIF by prolyl hydroxylases: recruitment of the candidate tumor suppressor protein ING4. Cell Cycle, 4: 1153-1156, 2005.
29. Ozer, A., Wu, L. C., and Bruick, R. K. The candidate tumor suppressor ING4 represses activation of the hypoxia inducible factor (HIF). Proc Natl Acad Sci U S A, 102: 7481-7486, 2005.
30. Cantley, L. C. The phosphoinositide 3-kinase pathway. Science, 296: 1655-1657, 2002.

31. Boronenkov, I. V., Loijens, J. C., Umeda, M., and Anderson, R. A. Phosphoinositide signaling pathways in nuclei are associated with nuclear speckles containing pre-mRNA processing factors. Mol Biol Cell, 9: 3547-3560, 1998.
32. Zhao, K., Wang, W., Rando, O. J., Xue, Y., Swiderek, K., Kuo, A., and Crabtree, G. R. Rapid and phosphoinositol-dependent binding of the SWI/SNF-like BAF complex to chromatin after T lymphocyte receptor signaling. Cell, 95: 625-636, 1998.
33. Gozani, O., Karuman, P., Jones, D. R., Ivanov, D., Cha, J., Lugovskoy, A. A., Baird, C. L., Zhu, H., Field, S. J., Lessnick, S. L., Villasenor, J., Mehrotra, B., Chen, J., Rao, V. R., Brugge, J. S., Ferguson, C. G., Payrastre, B., Myszka, D. G., Cantley, L. C., Wagner, G., Divecha, N., Prestwich, G. D., and Yuan, J. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell, 114: 99-111, 2003.
34. Shi, X. and Gozani, O. The fellowships of the INGs. J Cell Biochem, 2005.
35. Soliman, M. A. and Riabowol, K. After a decade of study-ING, a PHD for a versatile family of proteins. Trends Biochem Sci, 32: 509-519, 2007.
36. Kaadige, M. R. and Ayer, D. E. The polybasic region that follows the plant homeodomain zinc finger 1 of Pf1 is necessary and sufficient for specific phosphoinositide binding. J Biol Chem, 281: 28831-28836, 2006.
37. Gunduz, M., Nagatsuka, H., Demircan, K., Gunduz, E., Cengiz, B., Ouchida, M., Tsujigiwa, H., Yamachika, E., Fukushima, K., Beder, L., Hirohata, S., Ninomiya, Y., Nishizaki, K., Shimizu, K., and Nagai, N. Frequent deletion and down-regulation of ING4, a candidate tumor suppressor gene at 12p13, in head and neck squamous cell carcinomas. Gene, 356: 109-117, 2005.
38. Shiseki, M., Nagashima, M., Pedeux, R. M., Kitahama-Shiseki, M., Miura, K., Okamura, S., Onogi, H., Higashimoto, Y., Appella, E., Yokota, J., and Harris, C. C. p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity. Cancer Res, 63: 2373-2378., 2003.
39. Kim, S., Chin, K., Gray, J. W., and Bishop, J. M. A screen for genes that suppress loss of contact inhibition: identification of ING4 as a candidate tumor suppressor gene in human cancer. Proc Natl Acad Sci U S A, 101: 16251-16256, 2004.
40. Kim, S. HuntING4 new tumor suppressors. Cell Cycle, 4: 516-517, 2005.
41. Unoki, M., Shen, J. C., Zheng, Z. M., and Harris, C. C. Novel splice variants of ING4 and their possible roles in the regulation of cell growth and motility. J Biol Chem, 281: 34677-34686, 2006.
42. Li, X., Cai, L., Liang, M., Wang, Y., Yang, J., and Zhao, Y. ING4 induces cell growth inhibition in human lung adenocarcinoma A549 cells by means of Wnt-1/beta-catenin signaling pathway. Anat Rec (Hoboken), 291: 593-600, 2008.
43. Zhang, X., Xu, L. S., Wang, Z. Q., Wang, K. S., Li, N., Cheng, Z. H., Huang, S. Z., Wei, D. Z., and Han, Z. G. ING4 induces G2/M cell cycle arrest and enhances the chemosensitivity to DNA-damage agents in HepG2 cells. FEBS Lett, 570: 7-12, 2004.
44. Xie, Y., Zhang, H., Sheng, W., Xiang, J., Ye, Z., and Yang, J. Adenovirus-mediated ING4 expression suppresses lung carcinoma cell growth via induction of cell cycle alteration and apoptosis and inhibition of tumor invasion and angiogenesis. Cancer Lett, 271: 105-116, 2008.

45. Brat, D. J., Bellail, A. C., and Van Meir, E. G. The role of interleukin-8 and its receptors in gliomagenesis and tumoral angiogenesis. Neuro Oncol, 7: 122-133, 2005.
46. Semenza, G. L. Targeting HIF-1 for cancer therapy. Nat Rev Cancer, 3: 721-732, 2003.
47. Wenger, R. H. Cellular adaptation to hypoxia: O2-sensing protein hydroxylases, hypoxia-inducible transcription factors, and O2-regulated gene expression. Faseb J, 16: 1151-1162, 2002.
48. Couture, J. F., Collazo, E., and Trievel, R. C. Molecular recognition of histone H3 by the WD40 protein WDR5. Nat Struct Mol Biol, 13: 698-703, 2006.
49. Jenuwein, T. and Allis, C. D. Translating the histone code. Science, 293: 1074-1080, 2001.
50. Santos-Rosa, H. and Caldas, C. Chromatin modifier enzymes, the histone code and cancer. Eur J Cancer, 41: 2381-2402, 2005.
51. Fischle, W., Wang, Y., and Allis, C. D. Histone and chromatin cross-talk. Curr Opin Cell Biol, 15: 172-183, 2003.
52. Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev, 18: 1251-1262, 2004.
53. Turner, B. M. Decoding the nucleosome. Cell, 75: 5-8, 1993.
54. Lall, S. Primers on chromatin. Nat Struct Mol Biol, 14: 1110-1115, 2007.
55. Russell, M., Berardi, P., Gong, W., and Riabowol, K. Grow-ING, Age-ING and Die-ING: ING proteins link cancer, senescence and apoptosis. Exp Cell Res, 312: 951-961, 2006.
56. Skowyra, D., Zeremski, M., Neznanov, N., Li, M., Choi, Y., Uesugi, M., Hauser, C. A., Gu, W., Gudkov, A. V., and Qin, J. Differential association of products of alternative transcripts of the candidate tumor suppressor ING1 with the mSin3/HDAC1 transcriptional corepressor complex. J Biol Chem, 276: 8734-8739, 2001.
57. Avvakumov, N. and Cote, J. The MYST family of histone acetyltransferases and their intimate links to cancer. Oncogene, 26: 5395-5407, 2007.
58. Contzler, R., Regamey, A., Favre, B., Roger, T., Hohl, D., and Huber, M. Histone acetyltransferase HBO1 inhibits NF-kappaB activity by coactivator sequestration. Biochem Biophys Res Commun, 350: 208-213, 2006.
59. Sharma, M., Zarnegar, M., Li, X., Lim, B., and Sun, Z. Androgen receptor interacts with a novel MYST protein, HBO1. J Biol Chem, 275: 35200-35208, 2000.
60. Georgiakaki, M., Chabbert-Buffet, N., Dasen, B., Meduri, G., Wenk, S., Rajhi, L., Amazit, L., Chauchereau, A., Burger, C. W., Blok, L. J., Milgrom, E., Lombes, M., Guiochon-Mantel, A., and Loosfelt, H. Ligand-controlled interaction of histone acetyltransferase binding to ORC-1 (HBO1) with the N-terminal transactivating domain of progesterone receptor induces steroid receptor coactivator 1-dependent coactivation of transcription. Mol Endocrinol, 20: 2122-2140, 2006.
61. Yang, X. J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res, 32: 959-976, 2004.
62. Iizuka, M., Sarmento, O. F., Sekiya, T., Scrable, H., Allis, C. D., and Smith, M. M. Hbo1 Links p53-dependent stress signaling to DNA replication licensing. Mol Cell Biol, 28: 140-153, 2008.

63. Gill, G. SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? Genes Dev, 18: 2046-2059, 2004.
64. Nightingale, K. P., O'Neill, L. P., and Turner, B. M. Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. Curr Opin Genet Dev, 16: 125-136, 2006.
65. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. Active genes are tri-methylated at K4 of histone H3. Nature, 419: 407-411, 2002.
66. Ruthenburg, A. J., Allis, C. D., and Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell, 25: 15-30, 2007.
67. Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., and Schreiber, S. L. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci U S A, 99: 8695-8700, 2002.
68. Ng, H. H., Robert, F., Young, R. A., and Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell, 11: 709-719, 2003.
69. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol, 14: 1025-1040, 2007.
70. de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M. A. Do protein motifs read the histone code? Bioessays, 27: 164-175, 2005.
71. Ma, J. C. and Dougherty, D. A. The Cationminus signpi Interaction. Chem Rev, 97: 1303-1324, 1997.
72. Jacobs, S. A. and Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science, 295: 2080-2083, 2002.
73. Jacobs, S. A., Taverna, S. D., Zhang, Y., Briggs, S. D., Li, J., Eissenberg, J. C., Allis, C. D., and Khorasanizadeh, S. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. Embo J, 20: 5232-5241, 2001.
74. Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V., and Laue, E. D. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature, 416: 103-107, 2002.
75. Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S., and Waters, M. L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. Proc Natl Acad Sci U S A, 104: 11184-11188, 2007.
76. Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature, 438: 1181-1185, 2005.
77. Sims, R. J., 3rd, Chen, C. F., Santos-Rosa, H., Kouzarides, T., Patel, S. S., and Reinberg, D. Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J Biol Chem, 280: 41789-41792, 2005.
78. Aasland, R., Gibson, T. J., and Stewart, A. F. The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci, 20: 56-59, 1995.

79. Pascual, J., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. J Mol Biol, 304: 723-729., 2000.

80. Schindler, U., Beckmann, H., and Cashmore, A. R. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. Plant J, 4: 137-150, 1993.

81. Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bahler, J., Green, R. D., and Kouzarides, T. Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. Nature, 2007.

82. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature, 442: 100-103, 2006.

83. Champagne, K. S., Saksouk, N., Pena, P. V., Johnson, K., Ullah, M., Yang, X. J., Cote, J., and Kutateladze, T. G. The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. Proteins, 72: 1371-1376, 2008.

84. Pena, P. V., Hom, R. A., Hung, T., Lin, H., Kuo, A. J., Wong, R. P., Subach, O. M., Champagne, K. S., Zhao, R., Verkhusha, V. V., Li, G., Gozani, O., and Kutateladze, T. G. Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. J Mol Biol, 380: 303-312, 2008.

OBJETIVOS

Esta tesis tiene como objetivo general el estudio de la relación estructura-función de ING4, uno de los miembros de la familia de supresores tumorales ING, y los siguientes objetivos particulares:

1. **Estudio de la organización estructural de la proteína ING4.**

2. **Caracterización estructural de la interacción entre ING4 y p53.**

3. **Reconocimiento de fosfatidilinositol fosfatos por parte del PHD y el extremo C-terminal de ING4.**

4. **Caracterización del reconocimiento de colas modificadas de histonas por parte del dominio PHD de ING4.**

5. **Estudio de la interacción PHD-H3K4me3 en el contexto de la proteína ING4 completa y en un entorno de alta ocupación de macromoléculas que mimetice el interior celular.**

ESTRATEGIA Y ABORDAJE EXPERIMENTAL

La estrategia para el estudio de la relación estructura-función de ING4 se diseñó a partir de la asunción de que la proteína ING4 es una proteína multidominio, según se desprende del análisis de su secuencia de aminoácidos. En base a alineamientos de secuencia y predicciones de estructura se definieron los diferentes motivos y dominios de ING4, y se delimitaron tres regiones de la proteína: D1, D2 y D3, para su estudio en forma aislada o en combinación. Esto permitía relacionar los datos estructurales y biofísicos con los datos funcionales descritos para los diferentes motivos de secuencia conservada de ING4.

A lo largo del trabajo de tesis se usaron diferentes técnicas estructurales y biofísicas para abordar los diferentes tipos de caracterización, intentando utilizar en todo momento la técnica que pudiera dar más información de la forma más eficaz y sencilla posible.

Para la caracterización estructural de ING4 se han utilizado la resonancia magnética nuclear (RMN), la cristalografía de rayos X y la dispersión de rayos X a bajo ángulo (SAXS). La RMN permite analizar y obtener la estructura tridimensional de proteínas en disolución, mientras que la cristalografía de rayos X requiere un cristal de proteína, (los cuales contienen una gran proporción de solvente). Las medidas de RMN permiten, además, caracterizar en detalle ciertos aspectos de la dinámica interna de la proteína. Por su parte, la técnica de SAXS permite obtener la forma global de la molécula, una información estructural de muy baja resolución que resulta útil cuando no es posible una determinación estructural de alta resolución por cristalografía o RMN. No se ha intentado obtener imágenes de ING4 por microscopía electrónica debido al bajo contraste esperado para una proteína de su masa, muy pequeña para esta técnica, y más aún con dominios separados por regiones probablemente desordenadas o flexibles. Mientras que la aplicación de todo el potencial de la espectroscopía de RMN está limitada por el tamaño de la molécula y es necesario disponer de grandes cantidades de proteína con enriquecimiento isotópico en ^{15}N y/o ^{13}C, para la cristalografía el factor limitante es conseguir un monocristal de calidad que difracte los rayos X. A diferencia de estas dos técnicas, el SAXS no requiere marcaje isotópico ni cristales de proteína, aunque sí una concentración elevada de proteína (del orden de 10 mg/mL)

En la medida de interacciones, la RMN es una técnica muy sensible y eficaz, pudiendo determinarse si existe interacción o no con otra molécula, el sitio de unión de la misma e incluso la constante de afinidad de la interacción. La RMN también permite determinar la estructura de complejos, siendo en este caso una técnica más laboriosa en comparación con la cristalografía de rayos X. Si se dispone de cristales con buenas propiedades de difracción, la cristalografía es una técnica mucho más rápida y eficaz (más aún si se dispone de un modelo adecuado para obtener las fases por reemplazamiento molecular).

Las medidas de calorimetría isoterma de titulación permiten el análisis cuantitativo de la termodinámica del equilibrio de interacción proteína-ligando obteniéndose los valores de la energía libre, sus componentes entálpica y entrópica, así como la estequiometría de la reacción.

El estado de oligomerización de ING4 y de sus dominios se ha caracterizado de varias formas pero son especialmente informativas las medidas de ultracentrifugación analítica, que luego se han intentado validar en el interior de la célula viva mediante ensayos de coinmunoprecipitación y de complementación de proteína fluorescente.

Durante la exploración inicial de la organización estructural de ING4, se observó un comportamiento muy favorable del dominio PHD para su determinación estructural por RMN. Esto nos llevó a determinar su estructura en disolución y a caracterizar su unión a diferentes ligandos en primer lugar, continuándose después con la caracterización estructural del resto de dominios de ING4 y la proteína completa.

SOLUTION STRUCTURE AND NMR CHARACTERIZATION OF THE BINDING TO METHYLATED HISTONE TAILS OF THE PLANT HOMEODOMAIN FINGER OF THE TUMOUR SUPPRESSOR ING4

FEBS Letters 580 (2006) 6903–6908
Received 1 November 2006; revised 17 November 2006; accepted 19 November 2006
Alicia Palacios, Pascal Garcia, Daniel Padró, Eva López-Hernández, Irene Martín and Francisco J. Blanco
NMR group, Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain

Abstract

Plant homeodomain (PHD) fingers are frequently present in proteins involved in chromatin remodelling, and some of them bind to histones. The family of proteins Inhibitors of Growth (ING) contain a PHD finger that bind to histone-3 trimethylated at lysine 4, and those of ING1 and ING2 also act as nuclear phosphoinositide receptors. We have determined the structure of ING4 PHD, and characterised its binding to phosphoinositides and histone methylated tails. In contrast to ING2, ING4 is not a phosphoinositide receptor and binds with similar affinity to the different methylation states of histone-3 at lysine 4.

Key words: ING4, PHD, histone binding, phosphoinositide binding, tumour suppressor, chromatin remodelling

Abbreviations: HSQC, heteronuclear single quantum coherence; NMR, nuclear magnetic resonance; NOESY, nuclear overhauser enhancement spectroscopy; PHD, plant homeodomain; ING, inhibitor of growth; RMSD, root mean square deviation; WT, wild type; DTT, dithiothreitol; DSS, 2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt; PIP5, D-myo-Phosphatidylinositol 5-phosphate; UV, ultra violet; CSP, chemical shift perturbation.

1. Introduction

The ING family of tumour suppressors[1] consist of five homologous proteins which form stable complexes with other proteins involved in the regulation of chromatin acetylation[2]. N-terminal histone tail modification is a key mechanism of regulation of chromatin structure, and the pattern of histone modification around a gene affects its transcription[3]. Histone acetylation and methylation at lysines are the most common modifications, and are recognised by specific protein domains[4]. ING proteins contain a conserved C-terminal PHD finger[5], also present in many nuclear proteins involved in gene expression regulation and chromatin remodelling[6]. The PHD of p300 and ACF1 bind to nucleosome histones, and since both proteins contain also a bromodomain, which recognise acetylated lysines, they could form an integrated nucleosome recognition module[7,8]. PHD fingers could bind preferentially to methylated ones, as do chromodomains. This has been confirmed by the recent report that the PHD of ING proteins[9,10] and the PHD of NURF[11,12] bind to histone-3 trimethylated at lysine 4 (H3K4me3).

The PHD fingers of ING1 and ING2 are also nuclear receptors of phosphoinositides[13]. These phospholipids recruit proteins to the vicinity of the membranes regulating cell survival, growth and proliferation. Their interaction with the PHD could regulate the nuclear response to cellular stress[14].

Here we describe the solution structure of the PHD finger of ING4 and the characterization of its binding to phosphoinositides and histone methylated tails. The results are compared with those reported for ING2 and their functional implications are discussed.

2. Materials and methods.

2.1 Protein expression and purification.

The PHD finger of ING4 (residues 188-249 with an extra methionine at the N-terminus) was subcloned into the expression vector pET11d from a plasmid harbouring the synthetic gene of ING4 with codons optimized for expression in E. coli (Entelechon GmbH). PHD mutants were constructed with QuickChange (Stratagene).

Proteins were over-expressed in *E. coli* BL21(DE3) cells grown at 37 ºC in rich medium supplemented with 50 μM $ZnCl_2$ and harvested after 4 hours of induction with 0.5 mM isopropyl-beta-D-thiogalactopiranoside. Labeled proteins were produced in minimal medium with $^{15}NH_4Cl$ and [$^{13}C_6$]-glucose. After sonication and ultracentrifugation, proteins were found predominantly in the pellet, solubilised in 6 M urea and refolded by a 1:10 fold dilution into cold 20 mM Tris pH 8.0, 50 μM $ZnCl_2$. Purification by anion-exchange chromatography and gel filtration yielded proteins whose identity was confirmed by mass spectrometry. A small amount of wild type PHD was purified directly from the soluble fraction yielding identical 1D NMR spectrum as the refolded protein.

2.2 NMR spectroscopy and structure determination.

NMR experiments were recorded on Bruker AVANCE 600 (with cryoprobe) and 700 spectrometers at 298 K in 20 mM sodium phosphate pH 6.5, 50 mM NaCl, 1mM deuterated DTT and 9% or 100 % D2O. Some samples also contained 0.03% NaN_3. Backbone and sidechain resonance assignment were obtained using a set of triple resonance experiments recorded on a 1.2 mM PHD sample. Chemical shifts were measured relative to internal DSS for ^1H and calculated for ^{15}N and ^{13}C[15]. Spectra were processed with XWINMR (Bruker) or NMRPipe[16] and analyzed using NMRView[17]. Distance restraints were obtained from 2D-NOESY and 3D-NOESY spectra edited in ^{13}C or ^{15}N (120ms mixing time). Dihedral angle restraints were obtained from an HNHA spectrum and from the backbone chemical shifts using TALOS[18]. Structures were calculated with DYANA[19] and used for NOE assignment in an iterative manner. The structures were refined by energy minimization with AMBER 7.0[20] (see supplementary material for the table with the structure statistics for the ensemble of the 25 refined models). The resonance assignment has been deposited with the BMRB entry 710. The refined models have been deposited in the Protein Data Bank with the entry XXXX. These structures are similar to those deposited as PDB entries 1WEN and 1WEU, which contain long extraneous residues at the chain termini and do not include the last four residues of the ING4 sequence.

2.3 Ligand Binding .

Methylated histone peptides were purchased from NeoMPS, Strasbourg, and contain an extra tyrosine residue at the C-terminus to measure peptide concentration by UV absorbance. The sequences of the unmodified histone peptides are NH_2-ART<u>K</u>QTAR<u>K</u>STGGKAY-COOH (residues 1-15 of histone-3), and NH_2-GGAKRH<u>K</u>VLRDNIQY-COOH (residues 14-27 of histone-4). The residues that were methylated in the different peptides are underlined. Stock peptide solutions (5-6 mM) were prepared in 20 mM sodium phosphate pH 6.5, 50 mM NaCl, and the binding was identified by the perturbation in the chemical shifts observed in ^1H-^{15}N-HSQC spectra of 50 μM PHD samples in the absence or presence of a 1:4 excess peptide dialysed simultaneously against the same buffer. Titrations were performed by

stepwise addition of peptide stock solutions into 50 μM PHD samples and measuring the changes in the chemical shifts of W237 peak in ^{1}H-^{15}N-HSQC. Dissociation constants (K_D) were determined by data fitting (Origin, Microcal) to the equation: $\Delta\delta=(K_D+[P]+[L]-sqrt((K_D+[P]+[L])^2-4*[P]*[L]))/(2*[P])*\Delta\delta_{max}$, where [L] is the concentration of the peptide, [P] is the concentration of PHD, $\Delta\delta$ is the measured CSP, and $\Delta\delta_{max}$ is the maximum difference in chemical shifts of the free protein and the ligand-bound protein. $\Delta\delta$ was calculated from the equation: $\Delta\delta=sqrt(((\Delta\delta_H)^2+(\Delta\delta_N/5)^2)*0.5)$, where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift changes in the ^{1}H and ^{15}N resonances, respectively, upon peptide addition.
Soluble PI5P was from Echelon Biosciences. Its binding was tested with ^{1}H-^{15}N HSQC spectra of 60 μM PHD in the absence or presence of a 1:10 excess of PI5P.

3. Results and discussion

The PHD finger of ING4 requires Zn^{2+} for proper folding. An interleaved finger scaffold consisting of the C4HC3 sequence motif coordinates two Zn^{2+} atoms that stabilise the loops, the antiparallel β-sheet and the one turn helix (figure 1), which form the typical fold of this domain[6,21]. The structure is well defined with higher variability at the chain termini, which is due to increased mobility as shown by heteronuclear ^{1}H-^{15}N NOEs (data not shown). The surface of the molecule has a region with high density of positive charge (figure 1c) with a large contribution from the last four residues at the C-terminus (RKKK). This region could be involved in the binding to phosphoinositides, in a similar way as a positively charge region present in ING1 and ING2 after the PHD sequences (see supplementary

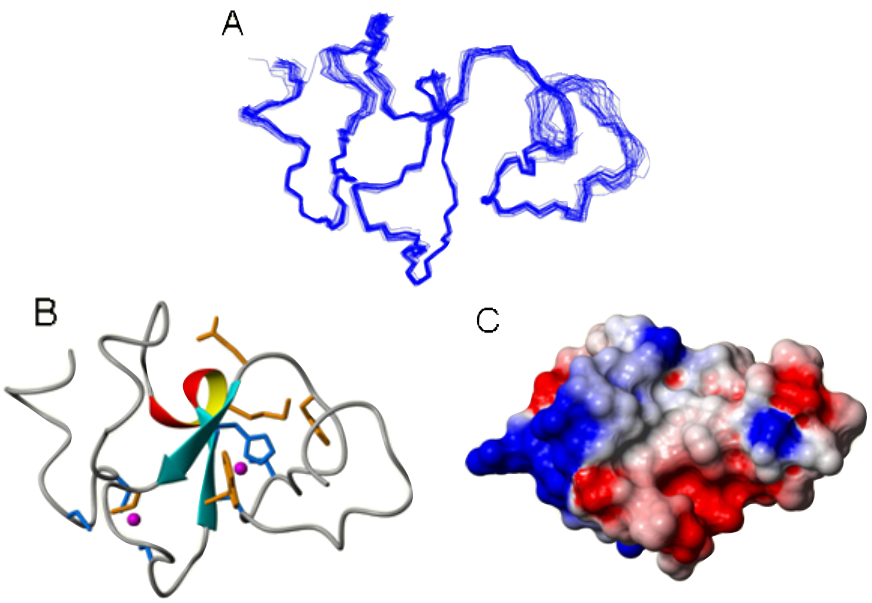

Figure 1. Solution structure of the ING4 PHD finger. (A) Ensemble of 25 refined structures; (B) ribbon model of one of the structures with the two Zn^{2+} ions in magenta, the side chains of the residues coordinating the ions in blue, and the side chains of the residues that experience the largest CSP upon binding to peptide H3K4me3 in orange (Y206, E208, M209, C212, W221, and G235 with CSP above the average plus one standard deviation); (C) Surface of the molecule coloured according to its electrostatic potential (negative in red and positive in blue). The three representations show the molecule in the same orientation.

material) are necessary for phosphoinositide binding[13]. Similar findings have been reported for the PHD of Pf1[22]. However, ING4 PHD finger does not bind to a panel of different phosphoinositides, or does so with an extremely low affinity, undetectable in solution by NMR (figure 3), and barely detectable in an overexposed dot blot with immobilized phosphoinositides (see supplementary material). Phosphoinositide binding is not a property of all PHD fingers[13], not even of those in ING proteins.

Recently it has been reported that the ING PHD fingers bind to H3K4me3[9]. The binding site on ING2 has been mapped by NMR and mutagenesis, and the three dimensional structure of the complex determined by crystallography[10]. We have confirmed by NMR that ING4 PHD finger binds to H3K4me3 peptide, and mapped the binding site (figures 2 and 3). The CSP measured in the presence of 1:4 excess of peptide is represented for each residue in figure 3. There are many residues that experience large perturbations in their chemical shifts, indicating a large interaction surface. There is a strong similarity with the pattern of changes experienced by ING2 PHD (see figure 2 of reference 10), suggesting that the binding site is similar in both proteins. As shown in figure 3 (see also supplementary material figure 2) the PHD finger of ING4 binds to histone-3 and to its six possible methylated variants at K4 or K9 with the same binding site, but it does not bind to histone-4 or its different methylation states in K20 (only for H4K20me3 a few residues show changes just above the experimental error). The titration curves obtained for histone-3 peptides are shown in figure 4, and the calculated dissociation constants are summarized in table 1. H3K4me3 peptide binds with a K_D=4.0±0.7 μM, close to the value of 7.9±2 μM measured previously[10] (by fluorescence, and possibly not exactly the same PHD chain length). Table 1 also contains the corresponding dissociation constants reported for ING2 PHD[10] for comparison. Both ING2 and ING4 bind to histone-3 methylated tails and not to histone-4, and bind to H3K4me3 with a similar affinity (considering the estimated errors). But beside these similarities there are remarkable differences. ING4 PHD is less selective than ING2 towards K4 versus K9 methylated H3. ING4 is also less selective towards the different methylation states of H3K4 than ING2. These measurements suggest a different mode of binding for the PHD. To better characterise the binding site and compare it to the mode of binding of ING2 we have analysed four alanine mutants of ING4 PHD. Y198A mutant corresponds to the Y215A mutation in ING2, which strongly decrease the affinity of ING2 for H3K4me3[10]. The other three mutations map the relevance for binding of the residue that experiences the largest CSP (M209) and two other residues that are in regions where there is a cluster of large CSP (figure 3) and that were not probed in the ING2 study[10]: D192, at the N-terminal region, and W237, which lines a hydrophobic pocket where the methyl groups of the peptide A1 and T2 residues are buried. Mutant W237A is unfolded, as indicated by the sharp and non-dispersed signals observed in its NMR spectra (data not shown), and the addition of H3K4me3 was not enough to displace the folding equilibrium towards an NMR detectable population of the folded state. The other three mutants show dispersed ^{1}H-^{15}N-HSQC spectra and the dissociation constants were measured with a signal that was tentatively assigned to W237 according to its chemical shifts, similar to those measured for the WT. As can be seen in table 2, the three mutants bind to H3K4me3 peptide with affinities that are not very different from those of the WT (reduced or increased by a factor of 3 or 4). These results indicate that M209 plays a minor role in the binding even though is the residue that suffers the largest CSP, probably because it is very close to the trimethylated lysine, as occurs in the ING2 complex[10]. The D192 mutation is in the N-terminal region, and the methylated lysine points towards this end in the structure of the ING2 complex. This explains the observed CSP, and the affinity measured suggests that this region

of the protein probably does not make a large contribution to the binding. This residue was not present in the shorter ING2 PHD[10]. Given the similar pattern of the CSP in both PHD fingers, we expected that mutant Y198A would show a very much reduced affinity for H3K4me3, as in ING2. On the contrary, the affinity is reduced but only by a factor of 4 while a three orders of magnitude reduction in the K_D was reported for ING2 PHD[10]. Although the explanation for this difference will only be possible after examination of the structure of the complex of ING4 PHD with the H3K4me3 peptide, this observation

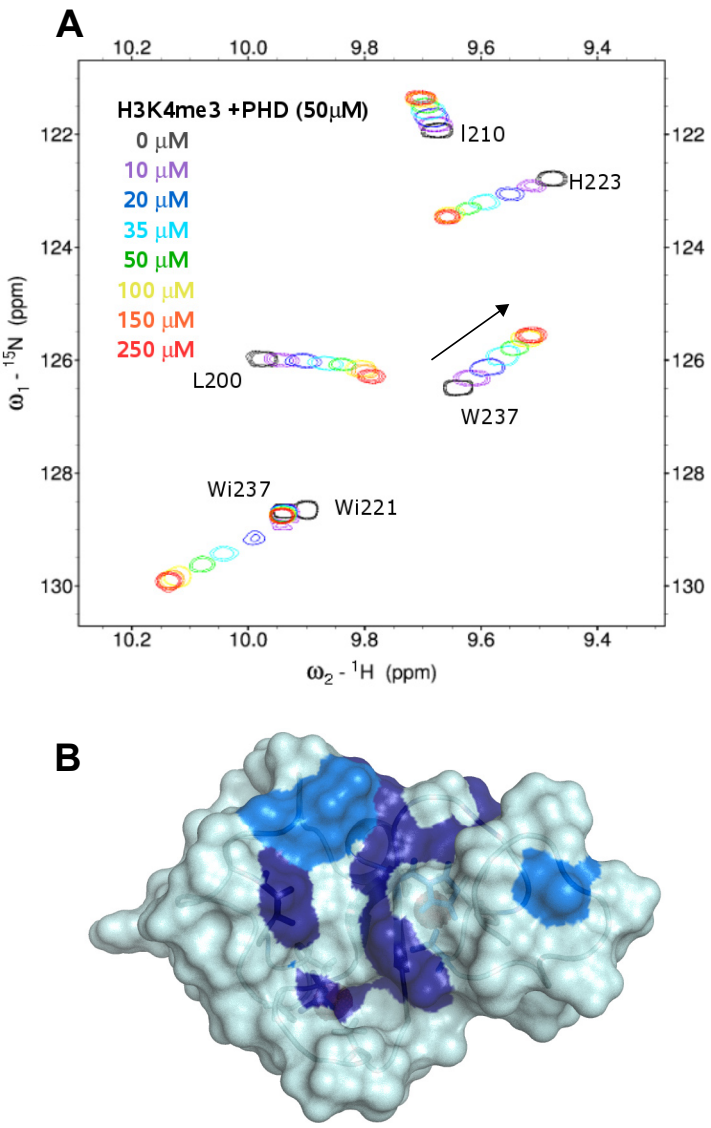

Figure 2. (A) Superposition of a region of six ^{1}H-^{15}N HSQC spectra of ING4 PHD after addition of different amounts of H3K4me3 peptide indicated with different colours. The labels adjacent to the signals indicate the corresponding residue ("Wi" stands for the triptophane indol NH signals). (B) Surface representation of the ING4 PHD finger (in pale blue with a ribbon model inside) with the same orientation as in figure 1. Those residues with a CSP upon binding to peptide H3K4me3 larger than the average plus one standard deviation (206, 208, 209, 212, 221, 235) are highlighted in dark blue, and those with a CSP larger than the average (192, 199, 204, 223, 232, 237) in light blue.

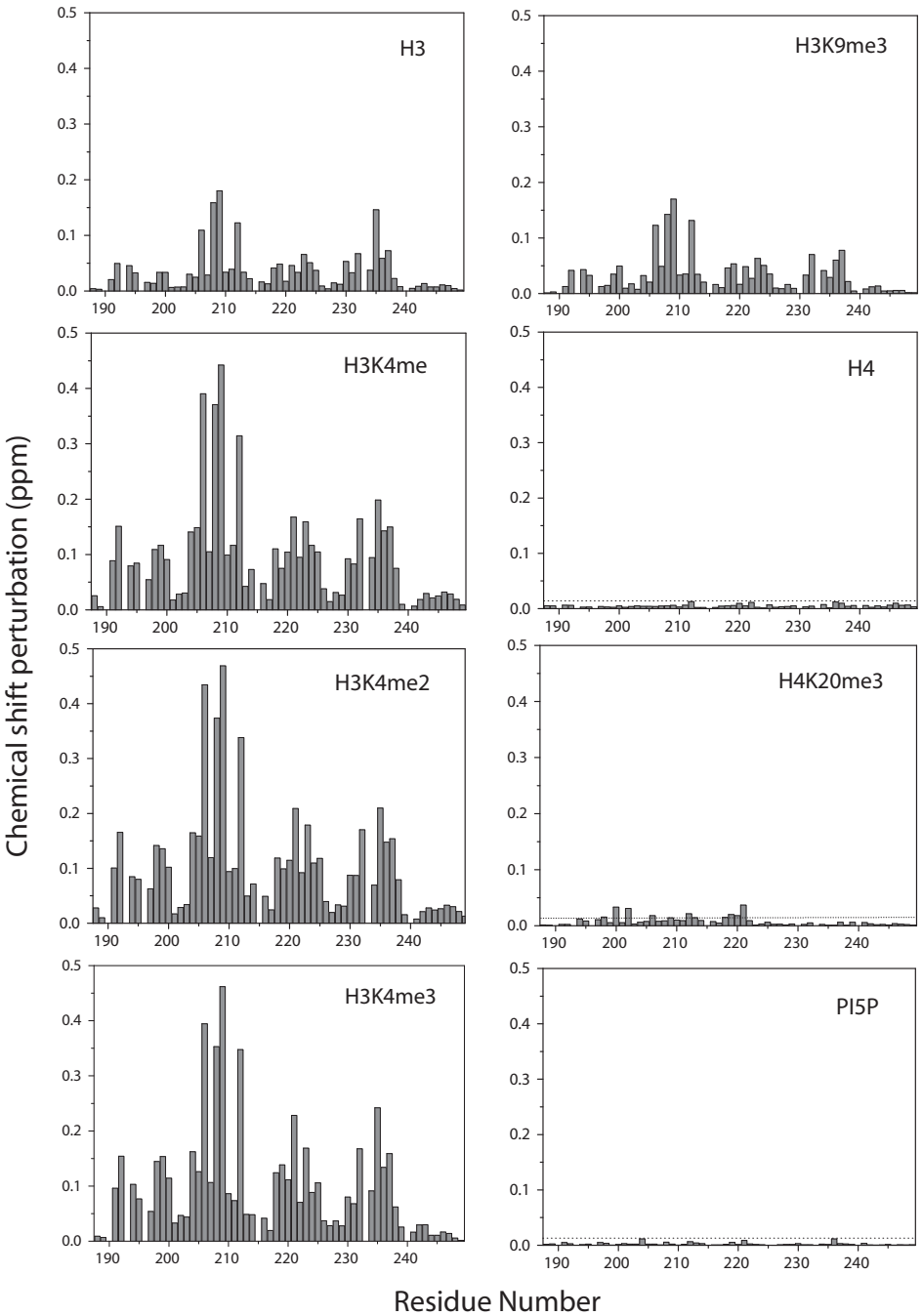

Figure 3. Binding histograms showing the CSP observed for each residue in the ^{1}H-^{15}N HSQC spectra of ING4 PHD in the presence of 1:4 excess peptide or 1:10 excess PIP5. The estimated experimental error (±0.012 ppm) is indicated by the dotted line in the two plots corresponding to peptide H4K20 peptides and PI5P binding.

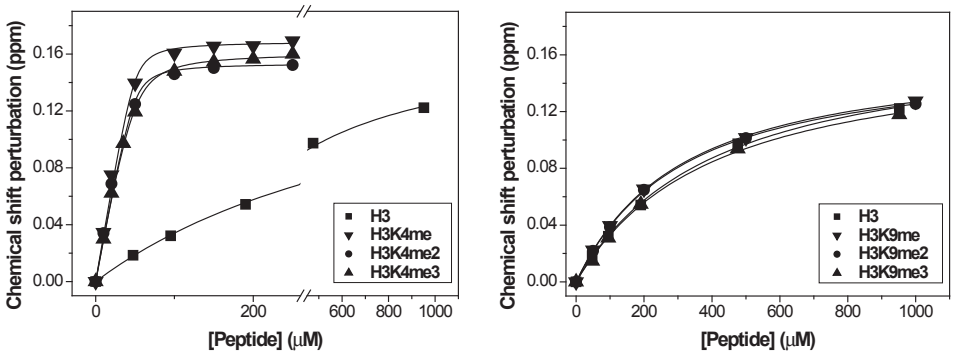

Figure 4. Plots of the CSP of W237 amide resonance of PHD (50 μM) as a function of the concentration of H3K4 (left), or H3K9 (right) peptides. The symbol´s height indicates the experimental error.

Table 1.

Dissociation constants for the binding to histone peptides of the PHD fingers of ING4 (this work) and ING2 (Ref. [10])

Ligand	K_D ING4 (μM)	K_D ING2 (μM)
H3	370 ± 20	2240 ± 350
H3K4me	1.6 ± 0.8	208 ± 80
H3K4me2	1.8 ± 1	15 ± 4
H3K4me3	4.0 ± 0.7	1.5 ± 1
H3K9me	274 ± 8	2380 ± 800
H3K9me2	274 ± 6	2320 ± 300
H3K9me3	360 ± 30	2000 ± 60
H4	›10000[a]	n.d.[b]
H4K20me	›10000[a]	›7000
H4K20me2	›10000[a]	›10000
H4K20me3	7900 ± 200[c]	›10000

[a]When no interaction was detected we assume a lower limit of 10000 μM.
[b]Not determined.
[c]Data up to 1:25 protein:peptide ratio were fitted assuming the same average $\Delta\delta_{max}$ obtained for the other peptides (0.16 mm.)

Table 2.

Dissociation constants for the mutants of ING4 PHD(this work) and ING2 (Ref. [10])

Mutant	K_D ING4 (μM)	K_D ING2 (μM)
D192A	0.8 ± 0.6	n.d.[a]
Y198A	12.0 ± 0.8	›500
M209A	17 ± 4	n.d.[a]
W237A	Unfolded	n.d.[a]

[a]Not determined.

demonstrates that there are important differences in the mode of binding for the two molecules. This is consistent with the different selectivity towards the different methylation states of histone-3 discussed above, which is higher for ING2. In this respect, ING4 behaves as the WDR5 module of the MLL1 complex, which activates transcription via methylation of histone-3[23]. WDR5 binds with similar affinities to the four H3K4 peptides (even the unmethylated one), and which is proposed to present the K4 chain of histone-3 for further methylation rather than read its methylation state.

The binding of ING2 PHD to H3K4me3 has been related with transcriptional repression through the recruitment of ING2-HDAC1 complex at target promoters[9,10], and ING2 has been copurified with mSin3A deacetylation complex, which is also linked to repression[2]. However, the PHD finger of NURF helps to recruit the NURF remodelling complex to promoters and modulate transcription initiation[11]. The binding properties of the PHD finger of ING4 link it with actively transcribed genes, since trimethylation of histone-3 at K4 is a hallmark of active genes[24], while trimethylation in K9 (and in K20 of histone-4) is associated with gene silencing[25]. ING4 has been found to copurify with histone-4 acetyl transferase complex HBO1, involved in transcription activation[2]. Although the precise functional implications of H3K4me3 recognition by the PHD finger of ING4 are still to be determined, the results presented here show that it behaves differently from ING2. While ING2 PHD performs as a dual specificity module for both H3K4me3 and phosphatidylinositol-5-phosphate[9] and plays a role in transcription repression, ING4 PHD does not bind to phosphoinositides, binds to the three methylation states of H3K4 and is possibly involved in transcription activation.

Acknowledgements

We thank David Pantoja and Tahl Zimmerman for help with scripts to work with nmrview and chemical shift data, Jaime Pascual for critical reading of the manuscript and Ramón Campos and Guillermo Montoya for helpful comments. This work was supported by grants GEN2003-28642 and BIO2003-02246 from the Spanish Ministry of Education and Science (MEC), a grant from Fundación de Investigación Médica Mutua Madrileña, and a MEC Ramón y Cajal contract from to FJB.

References

[1] Gong, W., Suzuki, K., Russell, M. and Riabowol, K. (2005) Function of the ING family of PHD proteins in cancer. Int J Biochem Cell Biol 37, 1054-65.

[2] Doyon, Y. et al. (2006) ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell 21, 51-64.

[3] Santos-Rosa, H. and Caldas, C. (2005) Chromatin modifier enzymes, the histone code and cancer. Eur J Cancer 41, 2381-402.

[4] de la Cruz, X., Lois, S., Sanchez-Molina, S. and Martinez-Balbas, M.A. (2005) Do protein motifs read the histone code? Bioessays 27, 164-75.

[5] Aasland, R., Gibson, T.J. and Stewart, A.F. (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci 20, 56-9.

[6] Bienz, M. (2006) The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci 31, 35-40.

[7] Eberharter, A., Vetter, I., Ferreira, R. and Becker, P.B. (2004) ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. Embo J 23, 4029-39.

[8] Ragvin, A. et al. (2004) Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. J Mol Biol 337, 773-88.

[9] Shi, X. et al. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature.

[10] Peña, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R. and Kutateladze, T.G. (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature.

[11] Wysocka, J. et al. (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature.

[12] Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D. and Patel, D.J. (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature.

[13] Gozani, O. et al. (2003) The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell 114, 99-111.

[14] Jones, D.R. et al. (2006) Nuclear PtdIns5P as a transducer of stress signaling: an in vivo role for PIP4Kbeta. Mol Cell 23, 685-95.

[15] Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 6, 135-40.

[16] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6, 277-93.

[17] Johnson, B.A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 278, 313-52.

[18] Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13, 289-302.

[19] Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273, 283-98.

[20] D.A. Case, D.A.P., J.W. Caldwell, T.E. Cheatham III, J. Wang, W.S. Ross, C.L., Simmerling, T.A.D., K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley,, V. Tsui, H.G., R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. and Singh, P.K.W.a.P.A.K. (2002) AMBER 7. University of California, San Francisco.

[21] Pascual, J., Martinez-Yamout, M., Dyson, H.J. and Wright, P.E. (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. J Mol Biol 304, 723-9.

[22] Shi, X. et al. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature 442, 96-9.

[23] Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J. and Verdine, G.L. (2006) Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat Struct Mol Biol 13, 704-12.

[24] Santos-Rosa, H. et al. (2002) Active genes are tri-methylated at K4 of histone H3. Nature 419, 407-11.

[25] Martin, C. and Zhang, Y. (2005) The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 6, 838-49.

Suplementary material

Methods.

For the phosphoinositide binding studies the gene of the PHD was inserted into plasmid pGEX-6P-2 (GE Healthcare) to obtain GST fusion proteins that were purified from the soluble fraction of LB grown cells by affinity GST chromatography (GE Healthcare GST-Trap columns) and gel filtration (Superdex 75, GE Healthcare).

PIP strips and soluble D-myo-Phosphatidylinositol 5-phosphate (PI5P) were from Echelon Biosciences. Incubation of GST-PHD molecules at 0.5 μg/ml with PIP strips was performed following the manufacturer recommended protocol. Binding was detected by a home made monoclonal anti-GST antibody and a secondary anti mouse IgG-HRP (Sigma) followed by ECL (GE Healthcare) detection.

Figures

Table 1.

Structural statistics for the ensemble of 25 refined NMR models of ING4 PHD.

Restraints	
All	1157
Distance	
Intraresidue	139
Short range	273
Medium range	225
Long range	461
Zinc coordination	8
Dihedral angles	
o	33
Ψ	18
χ_1	8
NOE violations	
Average	0.23 Å
Maximum	0.30 Å
Pairwise RMSD (all residues except initial Met)	
Backbone	0.7 ± 0.21
All heavy atoms	1.2 ± 0.17
Ramachandran map (all residues except initial Met)	
Residues in most favoured regions	58.0 %
Residues in additionally allowed regions	34.0 %
Residues in generously allowed regions	6.0 %
Residues in disallowed regions	2.0 %

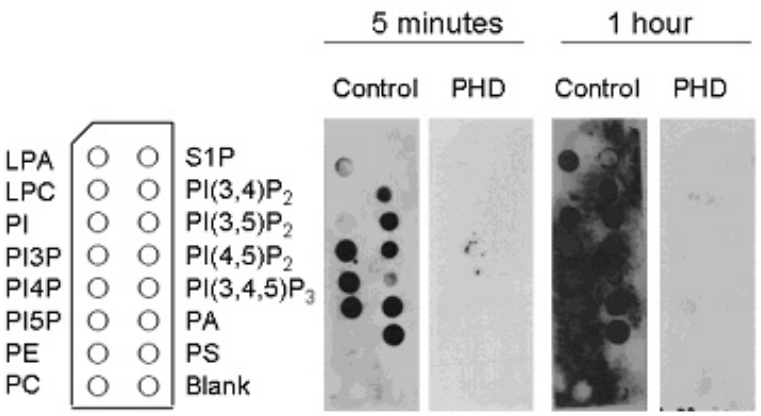

Figure 1. Binding of PHD molecules to membrane immobilized PIPs. From left to right: scheme of the lipids present on the PIP strip, blot with the control GST-fusion protein after 5 minutes exposure, blot with GST- PHD after 5 minutes exposure, blot with the control GST-fusion protein after 1 hour exposure, and blot with GST- PHD after 1 hour exposure.

```
ING1_PHD (202-279)   DLPIDPNEPTYCLCNQVSYGEMIGCDNDECPIEWFHFSCVGLNHKPKGKWYCPKCRGENEKTMDKALEKSKKERAYNR
ING2_PHD (204-280)   EFAIDPNEPTYCLCNQVSYGEMIGCDNEQCPIEWFHFSCVSLTYKPKGKWYCPKCRGDNEKTMDKSTEKTKKDRR-SR
ING3_PHD (352-418)   DWTYDPNEPRYCICNQVSYGEMVGCDNQDCPIEWFHYGCVGLTEAPKGKWYCPQCT---------AAMKRRGSRHK--
ING4_PHD (188-249)   DMPVDPNEPTYCLCHQVSYGEMIGCDNPDCSIEWFHFACVGLTTKPRGKWFCPRC----------SQERKKK------
ING5_PHD (178-240)   DMPVDPNEPTYCLCHQVSYGEMIGCDNPDCPIEWFHFACVDLTTKPKGKWFCPRC----------VQEKRKKK-----
```

Figure 2. Sequence alignment of the PHD sequences of the five ING human proteins. Black background indicate identical residues, and grey background conservative changes. The alignment was done with ClustalX (Thompson,J.D., Gibson,T. J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 24:4876-4882.)

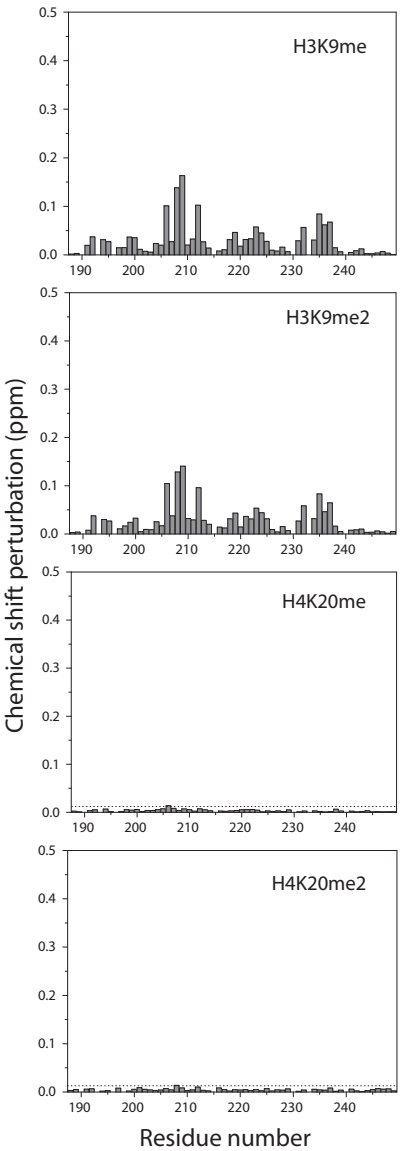

Figure 3. Binding histograms showing the chemical shift perturbations observed for each residue in the 1H-15N HSQC spectra of ING4 PHD finger in the presence of a 1:4 excess of peptides. The estimated experimental error (±0.012 ppm) is indicated by the dotted line in the two plots corresponding to peptides H4K20me and H4K20me2.

MOLECULAR BASIS OF HISTONE H3K4me3 RECOGNITION BY ING4

THE JOURNAL OF BIOLOGICAL CHEMISTRY
VOL. 283, NO. 23, pp. 15956–15964, June 6, 2008
Received December 10, 2007; revised March 18, 2008; published April 1, 2008
Alicia Palacios[$], Inés G. Muñoz[†], David Pantoja-Uceda[‡], María J. Marcaida[†], Daniel Torres[‡], José M. Martín-García[¶], Irene Luque[¶], Guillermo Montoya[†] and Francisco J. Blanco[$]

[$]*Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia, 48160 Derio, Spain; Structural Biology and Biocomputing Programme, [‡]NMR Group and [†]Macromolecular Crystallography Group, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández. Almagro 3, 28029-Madrid, Spain; and [¶]Department of Physical Chemistry and Institute of Biotechnology, Faculty of Sciences, University of Granada, 18071 Granada, Spain*

Abstract

The Inhibitors of Growth (ING) family of tumor suppressors consists of five homologous proteins involved in chromatin remodeling. They form part of different acetylation and deacetylation complexes, and are thought to direct them to specific regions of the chromatin, through the recognition of trimethylated-K4 in the histone-3 tail (H3K4me3) by their conserved Plant HomeoDomain (PHD). We have determined the crystal structure of ING4-PHD bound to H3K4me3, which reveals a tight complex stabilized by numerous interactions. NMR shows that there is a reduction in the backbone mobility on the regions of the PHD that participate in the peptide binding, and binding affinities differ depending on histone tail lengths Thermodynamic analysis reveals that the discrimination in favor of methylated lysine is entropy driven, contrary to what has been described for chromodomains. The molecular basis of H3K4me3 recognition by ING4 differs from that of ING2, which is consistent with their different affinities for methylated histone tails. These differences suggest a distinct role in transcriptional regulation for these two ING family members due to the antagonistic effect of the complexes that they recruit onto chromatin. Our results illustrate the versatility of PHD fingers as readers of the histone code.

Abbreviations: ING, inhibitor of growth; PHD, plant homeodomain; H3, histone 3; BPTF, bromodomain PHD finger transcription factor; NURF, nucleosome remodeling factor; ITC, Isothermal titration calorimetry; ASA, accessible surface area; CSP, chemical shift perturbations; HBO1, histone acetyltransferase binding to ORC 1; HP1, heterochromatin protein 1; CHD1, chromo-ATPase/helicase-DNA binding protein 1; H3YK4meX, N-terminal histone 3 tails of length Y with X methyls bound to the amino group of lysine 4; HDAC1, histone deacetylase complex 1; H3K9ac, histone 3 lysine 9 residue acetylated at its amino group; H3S10ph, histone 3 serine 10 residue phosphorylated at its hydroxyl group.

1. Introduction

Regulation of chromatin dynamics dictates the outcome of fundamental nuclear processes such as DNA transcription replication and repair (1-3). It is central to cell homeostasis, as alterations in chromatin structure contribute to the development of cancer and other human diseases (4). The ING family of tumor suppressors consists of five homologous proteins implicated in chromatin remodeling, growth arrest and, in cooperation with p53, in senescence and apoptosis (5-7). They are frequently deregulated in different types of cancer (8), and contain a conserved C-terminal PHD finger (9) which is present in many nuclear proteins involved in gen expression regulation and chromatin remodeling (10). They form stable histone acetylation or deacetylation complexes (11) and are thought to direct them to specific regions of the chromatin through binding of their PHD fingers to histone 3 N-terminal tails trimethylated at lysine 4 (12,13). These binding properties link ING proteins with actively transcribed genes, since H3K4 trimethylation is a hallmark of active genes (14). The recognition of H3K4me3 by ING2 is critical for the occupancy of the mSin3A-HDAC1 complex at the promoter of the cyclin D1 gene, which results in histone deacetylation and transcriptional repression of the active gene in response to DNA damage (15). This result suggests a general active transcriptional repression role for ING2; nonetheless, the biological outcome of the recognition of methylated histone tails by the other ING proteins is still unclear. Different

PHD fingers link H3K4me3 recognition with gene activation, such as the PHD of BPTF, which helps to recruit the NURF remodeling complex to target promoters modulating transcription initiation (16,17). Hence, the function of the PHD-H3K4me3 binding event, and its effect of transcription, is determined by the particular protein reader of this histone code mark.

We have solved the crystal structure of the PHD of ING4 bound to H3K4me3 tail, which remarkably shows a different mode of binding with respect to the previously reported for ING2, with a longer region of the histone tail participating in the interaction with ING4. The structure of the complex provides new insights into the determinants of the different binding affinities measured for the two domains (13), and allows for the interpretation of NMR data showing the stabilizing effect of C-terminal extensions of the histone peptide and a reduced backbone mobility in the PHD on peptide binding. Isothermal titration calorimetry measurements (ITC) show that the discrimination in favor of the methylated lysine is due to more favorable solvation entropy contributions.

2. Experimental Procedures

2.1 ING4 proteins and histone 3 peptides sample preparation

The clone of ING4 PHD finger 188-246 was made from the 188-249 construct using the QUickChange mutagenesis kit (Stratagene). Protein expression and purification was done as previously described (13). Synthetic lyophilized peptides were purchased from NeoMPS, Strasbourg, and correspond to histone-3 residues 1-10 (ARTKQTARKS) or residues 1-15 with an extra Tyr residue at the C-terminus (ARTKQTARKSTGGKAY) (13) with the four possible methylation states at lysine 4. The concentrations of the peptide stock solutions were measured by amino acid analysis (10 residue peptides) or by ultraviolet absorbance (using the absorbance of the aromatic side chain of the tyrosine residue in the 15+1 residue long peptides).

2.2 Crystallization and structure determination

Crystals of ING4(188-246) bound to $H3_{10}K4me3$ peptide were grown by the hanging drop vapor diffusion method at 5 °C. The pure protein was dialyzed against 10 mM Tris pH 6.5, 150 mM NaCl and 2 mM DTT and concentrated to 1.4 mM. The $H3_{10}K4me3$ peptide was used at a concentration of 35 mM in water. The complex was formed by mixing the PHD and the peptide in a 1:2 molar ratio. Crystallization drops were set up by mixing 1 µl of protein/peptide complex with 1 µl of reservoir solution of 35% PEG 6000 and 0.4 µl of 100 mM $CoCl_2$. Needles grew overnight but were not suitable for diffraction experiments. Good quality crystals appeared after using these needles for seeding in fresh drops that were incubated for 3 hours at 4°C. Crystals grew in about 3 days and reached final dimensions of 0.15 × 0.01 × 0.01 mm^3. Prior to data collection, protein complex crystals were immersed in the precipitant solution containing 5% (v/v) ethylene glycol, followed by rapid cooling in liquid nitrogen. A complete X-ray diffraction data set was collected at the beamline ID 29 (ESRF, Grenoble). Data was indexed, integrated and scaled with HLK2000 (18). Crystals belonged to the P43 space group with cell dimensions a = b = 68.51 Å, c = 28.51 Å, α, β, γ = 90º. Matthews coefficient and self rotation function indicated the presence of two molecules in the asymmetric unit, with a solvent content of 36.78%. Molecular replacement was performed with Phaser (19) using the PDB entry 2G6Q as model after removal of the

bound peptide. Refinement was carried out with REFMAC5 (20) including rigid-body refinement as the first step. Several rounds of iterative rebuilding with O (21) and refinement were performed. Two molecules of 1,4-dithiothreitol, and solvent water molecules were placed into the electron density using Arp/Warp (22). The final model contains 2 complexes in the asymmetric unit: the first constituted by ING4(195-244)/ H3(1-6) and the second, by ING4(191-244)/H3(1-10). Data collection details and statistics of the refinement can be found in Table 1. The coordinates have been deposited with the Protein Data Bank (accession code 2VNF).

2.3 NMR spectroscopy

NMR spectra were recorded at 25 °C in 20 mM sodium phosphate pH 6.5, 50 mM NaCl, 1mM dithiotreitol, 5% (v/v) $^{2}H_2O_2$, 0.01% Na_3N in a Bruker AVANCE 600 as described (13). Titrations were performed by stepwise addition of concentrated (5-6 mM) peptide stock solutions into 600 µL samples of 50 µM PHD. Dissociation constants were determined as described (13). The spectra to measure the differences in the CSP upon ING4 binding to $H3_{10}K4m3$ or $H3_{15}K4m3$ where obtained under identical conditions with two samples containing 50 µM PHD and a four fold excess of each peptide, which were simultaneously dialyzed against the same buffer. Backbone ^{15}N T_1, T_2 and $\{^{1}H\}$-^{15}N heteronuclear NOE measurements (23) were performed on a Bruker AVANCE 700 spectrometer on a 0.83 mM uniformly ^{15}N labeled PHD sample with or without 1.66 mM peptide. For free ING4-PHD ten time points (20, 60, 140, 240, 360, 460, 660, 860, 1100 and 1300ms) were collected for T_1 measurements and a different set of nine time points (16, 32, 63, 110, 158,190, 222, 270 and 396 ms) was collected to measure the T_2 values. T_1 and T_2 experiments were acquired with 8 scans and a repetition delay of 3 s, while each $\{^{1}H\}$-^{15}N NOE spectrum (both saturated and non-saturated one) was acquired with 136 scans and with an overall recycling delay of 10 s to ensure the maximal development of NOEs before acquisition and to allow solvent relaxation, thus avoiding transfer of saturation to the most exposed amide protons of the protein from scan to scan (24). Relaxation measurements for ING4-PHD bound to H3K4m3 peptides were carried out under the same conditions as those of the free form, but with one more sample time for T1 experiment (1600 ms) and different number of scans of the $\{^{1}H\}$-^{15}N NOE experiment 148 and 128 scans in the case of $H3_{10}K4m3$ and $H3_{15}K4m3$, respectively. In order to check the reproducibility of the relaxation measurements, the T_1 experiments for free ING4-PHD and the $\{^{1}H\}$-^{15}N NOE experiment for ING4-PHD bound to $H3_{15}K4m3$ were measured twice yielding the same results. Relaxation times were calculated via least squares fitting of peak intensities to a two parameter exponential function, using the rate analysis routine contained in the NMRView program (25). The heteronuclear NOEs were calculated from the ratio of cross-peak intensities in spectra collected with and without amide proton saturation during the recycle delay. Uncertainties in peak heights were determined from the standard deviation of the distribution of intensities in a region of the HSQC spectra where no signal and only noise was observed. The principal components of the inertia tensor were calculated with the Pdbinertia program (A.G. Palmer III, Columbia University) using the first model of the ensemble of 20 structures determined by NMR (Protein Data Bank entry 2JMQ), which is the best one according to restraint violations data, and the X-ray structure for the ING4-PHD/H3K4me3 complex. The estimation of the overall correlation time was obtained from the ratio of the mean values of T_1 and T_2, which were calculated from a subset of residues with little internal motion and no significant exchange broadening. This subset excluded those residues with NOEs smaller than 0.65 and also those residues with T_2 smaller than the average

minus one standard deviation, unless their corresponding T_1 values were larger than the average plus one standard deviation (26). The diffusion tensor, which describes the rotational diffusion anisotropy, was determined by two approaches (27,28) using the programs r2r1_diffusion and quadric_diffusion (A.G. Palmer III, Columbia University). The calculations failed when using the errors in T_1 and T_2 estimated by Monte Carlo simulations, which were unrealistically low. Therefore, the errors were scaled up by the minimum factor that allowed an interpretation of the data in terms of a rotational diffusion tensor. This procedure resulted in average errors of 9.5% and 10% for free and bound ING4-PHD, respectively. The ^{15}N relaxation was analyzed assuming dipolar coupling with the directly attached proton, with a bond length of 1.02 Å, and a contribution from the ^{15}N chemical shift anisotropy with a value of -160 ppm. The program FAST-Modelfree (29) which interfaces with the program MODELFREE version 4.2 (30) was used to fit the relaxation data to the model free formalism of Lipari and Szabo (31). Five different models of internal motion were evaluated for each amide ^{1}H-^{15}N pair: (i) S^2, (ii) S^2 and τ_e, (iii) S^2 and R_{ex}, (iv) S^2, τ_e, and R_{ex}, and (v) S_f^2, S^2, and τ_e, where S^2 is the generalized order parameter, τ_e is the effective internal correlation time, R_{ex} is the exchange contribution to the transverse relaxation and S_f^2 is related to the amplitude of the fast internal motions.

2.4 Isothermal titration calorimetry

ITC experiments were performed using a high-precision MCS titration calorimetric system (Microcal Inc., Northampton, MA). The ING4-PHD domain was extensively dialyzed against the titration buffer. All solutions were filtered, properly degassed to avoid bubble formation, and equilibrated to 25 °C prior to each experiment. The protein solution (at 40-65 μM) in the calorimetric cell was titrated with the appropriate ligand (at 600-800 μM) dissolved in the dialysis buffer following a profile of injection volumes from 2.8 μL to 20 μL to better define the titration curve. The heat evolved after each peptide injection was obtained from the integral of the calorimetric signal. The heat produced by the binding reaction between the PHD and the peptides was obtained as the difference between the heat of reaction and the corresponding heat of dilution, as obtained from independent titrations of the peptides into the buffer. The resulting binding isotherms were analyzed by non-linear least-square fittings of the experimental data to a model corresponding to a single set of identical sites, as described in supplemental data. For the interactions of $H3_{10}K4me0$ with ING4-PHD, for which the dissociation constant is out of the range measurable directly by ITC, displacement experiments using $H3_{10}K4me1$ as competing ligand were carried out. Briefly, a 65 μM ING4-PHD solution with $H3_{10}K4me0$ at a 1:4.6 molar ratio was placed in the calorimetric cell and titrated with $H3_{10}K4me1$ (at 2 mM) following a profile of injection volumes from 4 μL to 20 μL. The resulting binding isotherms, corrected for the dilution heats, were analyzed by non-linear least-square fittings of the experimental data to the exact displacement model as described (32) using the binding affinity and binding enthalpy for $H3_{10}K4me1$ obtained from the titrations with this peptide using the same injection profile. The data analysis was done with Microcal Origin (OriginLab Corporation, Northampton, MA) together with software developed in our laboratory.

2.5 Solvation energy calculations

The differences in solvation entropy were calculated according to Freire´s structural parameterization of the energetics (33) were calculated using the crystal structure of the ING4-PHD/$H3_{10}K4me3$ complex and two modeled structures for the di- and mono-methylated species. Of the three methyl groups in the

crystal structure, one is fully buried at the binding interface (C atom no. 891, ΔASA = 0 Å^2), the second is somewhat exposed (C atom no. 893, ΔASA = 16.93 Å^2) and the third one is significantly exposed (C atom no. 892, ΔASA = 26.93 Å^2). One could reasonably assume that in the dimethylated complex both methyl groups would tend to be as buried as possible. This is what is observed in the complex of the HP1 chromodomain with H3$_{15}$K9me2 and me3 peptides, where the methylated lysine side chains adopt a structure very similar to that in ING4-PHD/H3$_{10}$K4me3 complex (34). For this reason, in the model of H3$_{10}$K4me2 bound to ING4 the methyl group that is exposed the most (C atom no. 892) was removed from the structure of H3$_{10}$K4me3 while C atoms no. 892 and 893 were removed to build the model for the H3$_{10}$K4me1 complex.

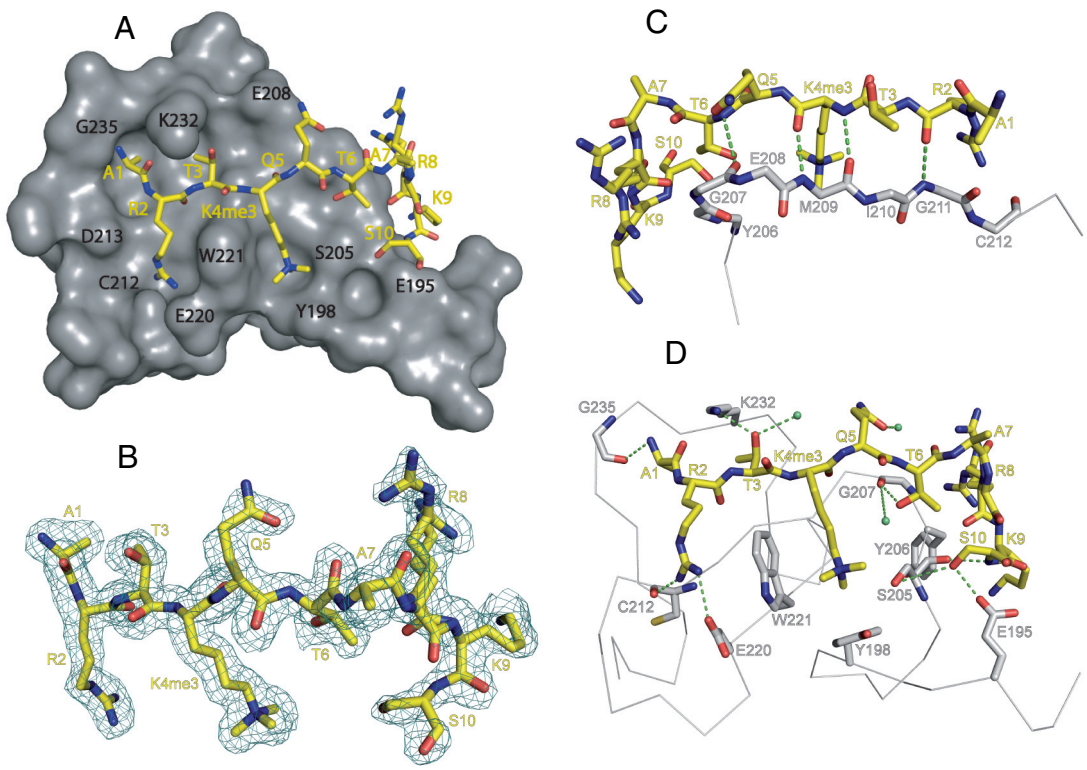

Figure 1. Three dimensional crystal structure of the complex of ING4-PHD bound to H3$_{10}$K4me3. A, overall structure with ING4-PHD shown in grey and H3$_{10}$K4me3 shown in yellow stick representation. The location of the ING4 residues lining the binding grooves for histone residues H3A1, R2, T3, K4, and S10 is indicated with black labels. B, view of a 2(|Fo| – |Fc|) omit map at 1.76 Å contoured at 1σ as a blue mesh. The omit map was calculated with the program OMIT in the CCP4 package (45). C, details of the β-sheet structure formed by residues H3$_{10}$R2-T6 and ING4 residues G207-G211 with backbone hydrogen bonds in green. D, Details of side chain interactions with hydrogen bonds and water molecules in green.

3. Results and Discussion

3.1 Recognition of H3K4me3 by ING4

Table 1.

X-Ray data collection and refinement statistics

Data collection[a]	
Environment	ADSC detector, ESRF, beamline ID29
Wavelength	1.072 Å
Cell dimensions(Å, °)	a=68.51, b=68.51, c=28.51, α=β=γ=90°
Resolution (Å)	68.51-1.76 (1.82-1.76)
Space Group	P43
Unique reflections	12817
Average multiplicity	3.8 (2.6)
Completeness (%)	96.9 (86.8)
Rmergeb	0.075 (0.38)
<I/σI)>	9.9 (1.7)
Refinement	
Number of reflections (completeness, %)	12178 (96.92)
Resolution range (Å)	68.51-1.76
R-factor / R-free (%)	15.62 / 22.67
Number of protein atoms (Average B, Å2)c	929 (30.61)
Number of water molecules (Average B, Å2)c	129 (38.31)
r.m.s bond length (Å)	0.012
r.m.s. bond angle (°)	1.816
Ramachandran plot outliers (number, %)d	3, 2.5%

[a] Values in the highest resolution shell are given in parentheses.
[b] $R_{merge} = \Sigma_\eta \Sigma_\iota | I_{\eta,\iota} - \langle I_\eta \rangle | / \Sigma_\eta \Sigma_\iota | I_{\eta,\iota}|$
[c]Calculated using MOLEMAN (43)
[d]Calculated with PROCHECK (44)

The crystal structure of ING4-PHD (residues 188–246) bound to histone 3 (residues 1–10) trimethylated at lysine4 ($H3_{10}$K4me3) was solved at 1.76 Å resolution (Fig 1A). The ten residues of the peptide were observed in the electron density, including two alternate conformations for H3R8 that could be modeled without ambiguity into their corresponding densities (Fig 1B). The peptide conformation does not seem to be influenced by the neighboring crystallographically related molecules, and the interactions with ING4 observed in the crystal structure are consistent with solution NMR data (Fig 2A). The structure of the complex shows that the N-terminal half of the histone tail binds to the surface of the PHD finger as a third strand of the anti-parallel β-sheet that forms the core of the PHD (35). The peptide N-terminus forms a hydrogen bond with the carbonyl of ING4 residue G235 and residues H3R2–T6, which have β-sheet backbone dihedral angles, form backbone hydrogen bonds with C212–G207 (Fig 1C). A kink in the main chain at residues H3A7,R8 recovers the extended conformation for residues H3K9,S10, with H3S10 forming a backbone hydrogen bond with S205. Due to the polar nature of histone-3 tail residues, the interactions are predominantly polar. The side chains of residues H3R2,T3,T6,S10 form salt bridges and hydrogen bonds with E220 and C212, K232, G207, and E195, respectively (Fig 1D), and the guanidinium group of H3R2 and the trimethylammonium group of H3K4me3 make cation-π interactions with the side chain of W221, which sits between the two basic peptide side chains. H3K4me3 forms another cation-π contact with Y198, which is not essential for recognition since the mutant Y198A still binds the peptide with moderate affinity (13). Residue H3K9 makes a cation-π interaction with Y206,

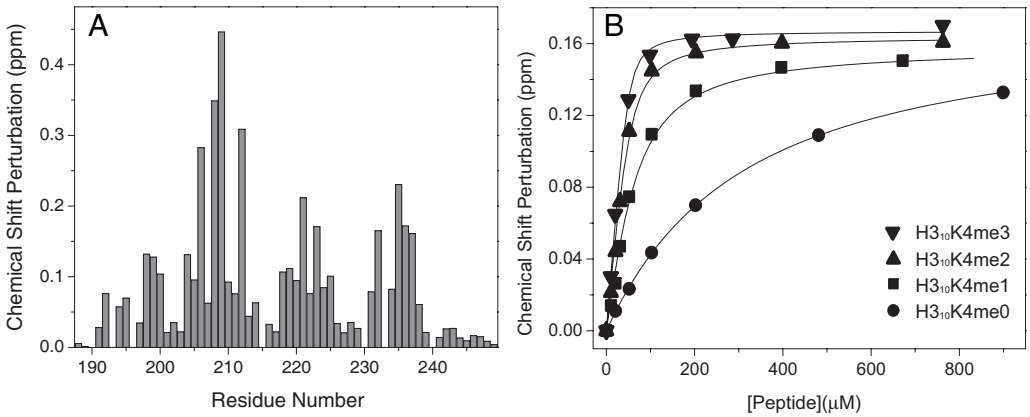

Figure 2. NMR analysis of the binding of ING4-PHD to $H3_{10}K4meX$ peptides. A, Binding histogram showing the CSP observed for each residue in the ^{1}H-^{15}N HSQC spectra of ING4 PHD in the presence of 1:4 excess of $H3_{10}K4me3$. B, the CSP of the W237 amide resonance of ING4(188-249) in ^{1}H-^{15}N-HSQC spectra is represented as a function of peptide concentration. The continuous line is the fitting to the binding model (13). The height of the symbol indicates the experimental error.

and so does H3R8 (at least in one of the two alternative conformations. However, these two interactions are probably weaker because while the distances between the charges and the aromatic rings are within the range of typical Van der Waals interactions, their relative orientations are not optimal- off the C_6 ring axis.

The PHD structure resembles an oblate ellipsoid with three grooves or channels in one side which are filled by the backbone and the side chains of residues H3A1-T6 (see supplemental Fig 1). The kink at H3A7 directs the peptide chain along the rim of the ellipsoid and H3S10 occupies a small depression lined by E195. The shape complementarities between the PHD and the bound peptide structures results in the burial of a large accessible surface area upon complex formation both in the PHD and in the peptide molecules (ASA, 1318 Å^2, see below), with H3A7 at the kink being the only residue not contributing to the buried ASA.

In the complex with ING4, the methylammonium of H3K4me3 is positioned in a cage formed by two aromatic residues (Y198 and W221), the hydrophobic side chain of M209, and S205. This binding mode of the methylated lysine is common to ING2, and similar to BPTF, with a cage of four aromatic residues, (17). Overall, the recognition of the N-terminal half of $H3_{10}K4me3$ by ING4 is very similar to that observed in the complex with ING2 (12), but differs markedly in the C-terminal half (Fig 3 and supplemental Fig 2). The peptide bound to ING2 is not kinked at H3A7 and the last two residues observed in the crystal (H3A7,R8) wander off the PHD. The different mode of binding of ING4, is consistent with the different effect of a homologous residue substitution (Y198A in ING4 and Y215A in ING2), which causes a three-fold reduction in the affinity of ING4 for $H3_{15}K4me3$ while it strongly destabilizes (more than three thousand-fold less affinity) the complex with ING2 (13). The reason for the different effect of the homologous mutation could be the distinct distribution of charges on the surface of the PHD fingers of ING2 and ING4. While in ING4 the N-terminal end of the molecule is predominantly negatively charged on the side closer to K4me3, it is positively charged in ING2 (Fig 3B). The shielding of the aromatic side chain of Y215 in ING2 may reduce the charge repulsion between this region and K4me3, and its elimination destabilizes the complex beyond the already destabilizing effect of removing

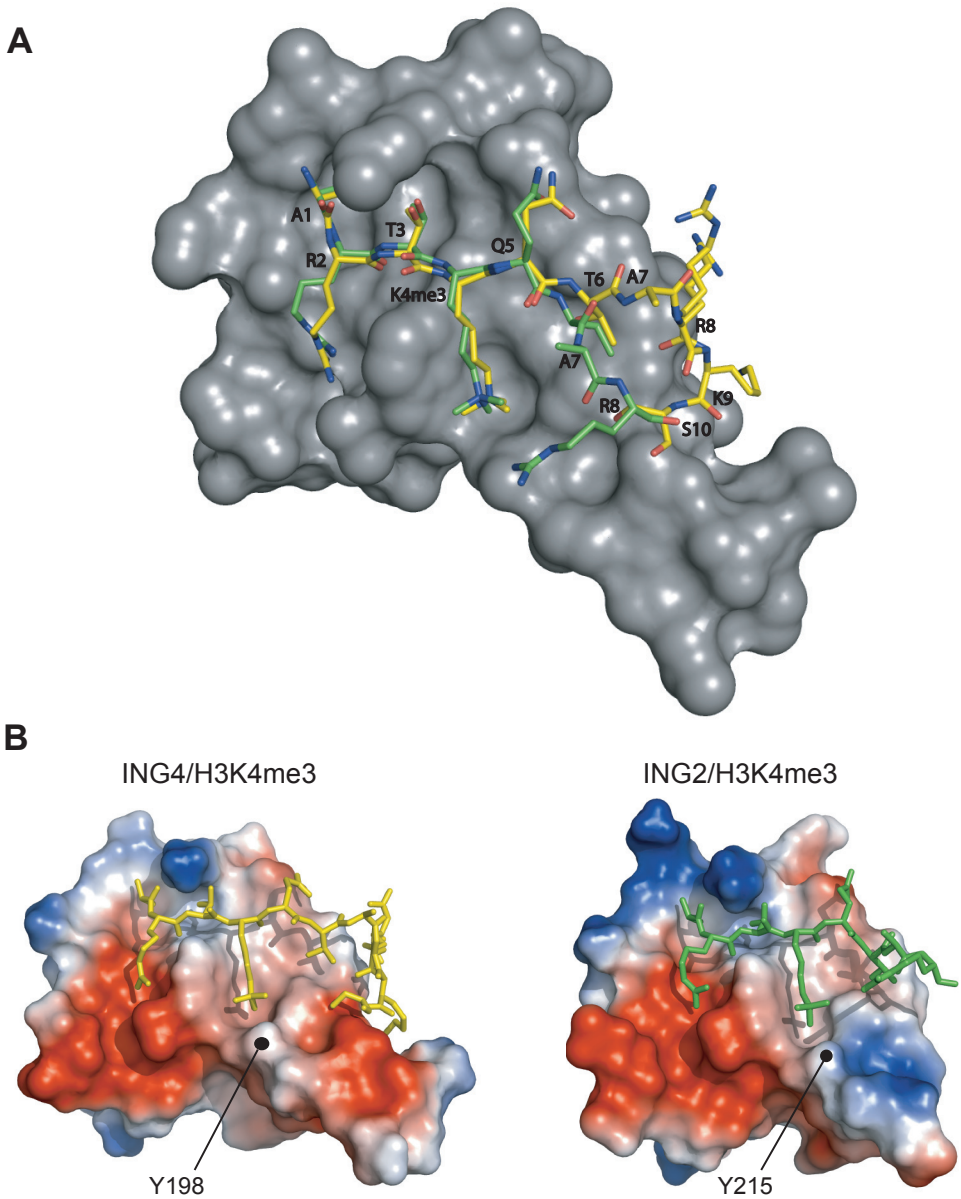

Figure 3. Comparison of the structures of H3K4me3 peptides bound to the ING4 and ING2 PHD. A, The crystal structure of ING4-PHD is shown as a grey surface with the bound $H3_{10}$K4me3 in yellow and the superimposed $H3_{12}$K4me3 from the corresponding ING2-PHD complex structure in green. Only residues H3A1 to H3R8 are seen in the crystal of the ING2 complex (Protein Data Bank entry 2G6Q). Black labels indicate the Cα atoms of the peptide residues, whose positions are very similar in both peptides up to H3T6. B, Surface representations of the PHD fingers of ING2 and ING4 bound to the histone peptides where the electrostatic potential is indicated by a gradient of red (negative charge) and blue (positive charge) colors. The position of the homologous tyrosine residues 198 and 215 is indicated in ING4 and ING2, respectively.

the favorable cation-π interaction. In the complex with ING4, the removal of the aromatic side chain of Y198 eliminates that same cation-π interaction, but it does not result in unfavorable electrostatic interactions.

3.2 ING4 binding to histone tails of different lengths

The recognition of the H3A1 by G235 blocks N-terminal extensions of the peptide chain bound to ING4, which is probably important for the specificity of the binding of $H3_{10}K4me3$ since the lysine modification is near the N-terminus (36). No similar limitation exists at the C-terminus and the extent of the interactions with other regions of the histone tail is unclear.

We have previously characterized the binding of ING4 to 15 residue long $H3_{15}K4meX$ peptides (13). The affinity for $H3_{10}K4me3$ is, within error, the same as for $H3_{15}K4me3$. Nonetheless, while ING4 does not discriminate between $H3_{15}K4me1$, -me2 or -me3, it does so for the shorter peptides, with the affinity increasing with the higher number of methyls (Fig 2B and Table 2), indicating that the additional residues contribute to the interaction with ING4. Crystallization trials with this longer peptide were unsuccessful, and the differences in the binding of $H3_{15}K4me3$ and $H3_{10}K4me3$ were examined in solution by NMR measuring the differences in the chemical shift perturbations (CSP) caused on the ING4 signals on binding to the two peptides. These measurements are very sensitive to weak interactions or those involving flexible regions of the ligand and/or receptor. The spectra of ING4 bound to $H3_{10}K4me3$ or $H3_{15}K4me3$ (Fig 4A) show several residues with larger perturbations caused by the longer peptide. Moreover, these perturbations are clustered in a region of ING4 in the vicinity of the C-terminal end of the peptide (Fig 4B) indicating that additional interactions take place, resulting in higher affinities for $H3_{15}K4me1$ and –me2 (Table 2). This finding is consistent with the decreased flexibility of the N-terminal end of ING4-PHD bound to $H3_{15}K4me3$ (see below).

Table 2.

Dissociation constants (K_D) measured in solution by NMR for the 10 and 15 residue long histone tails methylated at K4 bound to ING4(188-249)

Ligand[a]	K_D (μM) [b]
$H3_{15}K4me0$	370 ± 20
$H3_{10}K4me0$	274 ± 6
$H3_{15}K4me1$	1.6 ± 0.8
$H3_{10}K4me1$	34 ± 4
$H3_{15}K4me2$	2 ± 1
$H3_{10}K4me2$	9.2 ± 1.4
$H3_{15}K4me3$	3.9 ± 0.7
$H3_{10}K4me3$	3.0 ± 0.6
$H3_{10}R2me2K4me3$	19.2 ± 1.7
$H3_{10}R2me2$[c]	1400 ± 40

[a]The values for $H3_{15}K4$ peptides are from (13)

[b]The fitting errors are indicated.

[c]Measured from the CSP of C212 signal instead of W237 (see supplemental Fig 4).

3.3 Dynamics and thermodynamics of the recognition of H3K4me3 by ING4

The chain termini of ING4 are highly flexible in solution, as indicated by the small values of the order parameters (S^2) measured for the backbone ^{15}N atoms (Fig 5A), which are sensitive to movements of the N-H bond in ps to ns time scales. Residue G235 also shows increased mobility compared with

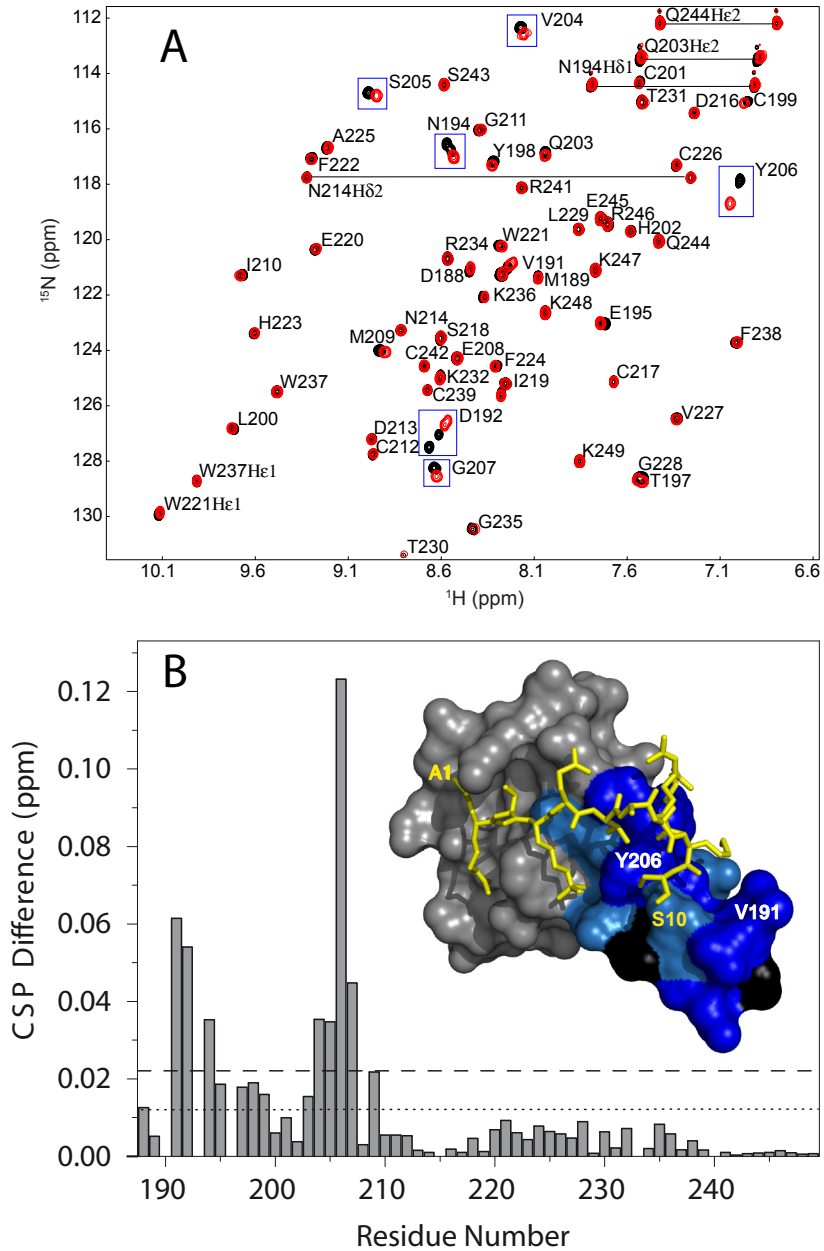

Figure 4. Binding of ING4-PHD to H3K4me3 peptides of different lengths. A, Overlay of HSQC spectra of ING4(188-249) bound to $H3_{15}$K4me3 (black) or $H3_{10}$K4me3 (red) Cross-peaks are labeled with their corresponding residue number and single letter code. Peaks from amide side chains are connected by straight lines. Residues with CSP differences larger than twice the experimental error are boxed. Four signals from a minor conformer due to the cis-trans isomerization of the peptidyl-prolyl bonds in the flexible N-terminal region of ING4 are labeled (M189c, V191c, D192c and N194c). A signal from an unknown molecule is marked with an asterisk. B, Binding histogram of the differences in the CSP for each ING4 PHD residue; the experimental error (± 0.012 ppm) is indicated by the dotted line. The inset is a surface representation of ING4 PHD where residues with 0 > ΔCSP > error, or ΔCSP > 2 × error are colored in light or dark blue, respectively. Proline residues, for which no signal can be observed, are colored in black.

the rest of the chain. Binding to $H3_{10}K4me3$ increases the order parameters at several regions of the ING4 chain containing residues directly involved in histone recognition (E195, G211, E220, G235). The increased order around E220 on peptide binding is consistent with the conformational exchange detected for this residue in free ING4 but not in the complex (Fig 5B). Conformational exchange in free ING4 is also detected for W221, which interacts with H3R2 and H3K4. These results show that the recognition of $H3_{10}K4me3$ is not made by a static PHD molecule, but that binding involves regions of ING4 with different degrees of mobility in its free form which become more rigid upon binding to the peptide. Relaxation data on ING4 bound to $H3_{15}K4me3$ show that the flexibility at the N-terminal region is further reduced on binding to this longer peptide (supplemental Fig 3), which is consistent with the observed contribution to the binding of the longer histone tails. The NMR spectra also show that the free peptides are highly flexible in solution and conformational order is induced on binding to ING4 (data not shown). A detailed account of the analysis of the backbone ^{15}N relaxation data can be found in the supplemental data.

The energetics of the ING4 binding to $H3_{10}K4meX$ was characterized by isothermal titration calorimetry. Low to moderate binding affinities, with a large increase when the first methyl is introduced were observed, in good agreement with those measured by NMR (Fig 6). The interaction is driven by a markedly exothermic binding enthalpy partially opposed by unfavorable entropic contributions

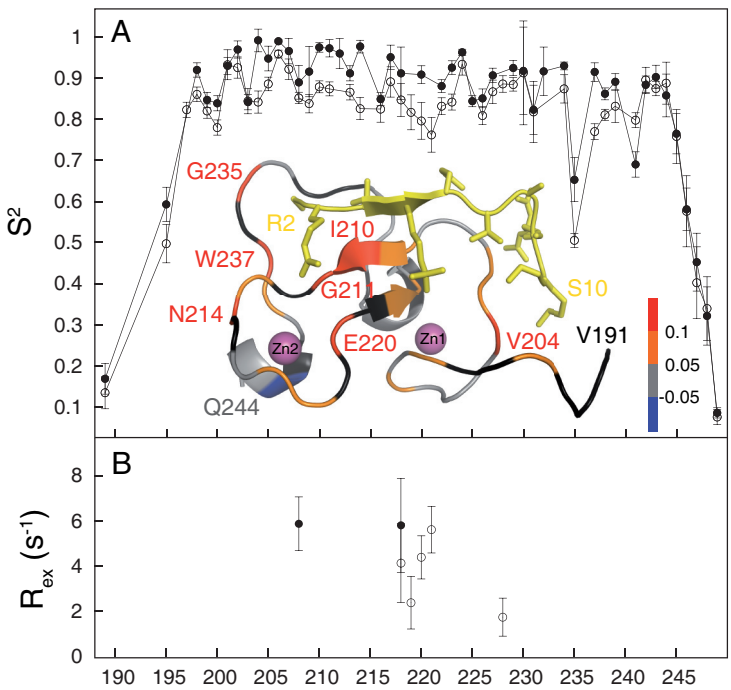

Figure 5. Backbone dynamics of ING4-PHD backbone.A, Backbone ^{15}N nuclei order parameters (S^2,) and B, the contribution of exchange processes (R_{ex}, bottom panel) to the relaxation of backbone ^{15}N nuclei are represented versus the sequence of ING4, in its free form (open circles) and in the complex with $H3_{10}K4me3$ (filled circles). The ribbon diagram shows the crystal structure of ING4-PHD bound to $H3_{10}K4me3$ (in yellow and with side chains). A color code indicates the change in S^2 for the corresponding ING4 residue on peptide binding. Residues for which the difference could not be measured are colored in black.

(Fig 7). This thermodynamic signature is consistent with the high density of hydrogen bonds and polar interactions at the binding interface, and with the burial of apolar and polar ASA upon complex formation (798 Å^2 and 520 Å^2 respectively, with a ratio $\Delta ASA_{apolar}/\Delta ASA_{polar}$ = 1.5). Nonetheless, the methylation state of the lysine residue does not have a significant effect on the enthalpic contributions to the binding affinity, which is very similar for the four peptides, so that the increment in the binding affinity upon methylation is entirely due to changes in the entropic contributions (Fig 7). It is interesting to note that the opposite situation has been proposed for chromodomains (37). Lysine methylation leads to a polarization of the C_ε-N_ζ bond increasing the cationic character of the methylammonium group and strengthening the cation-π interactions with the aromatic cage, but also results in an increment on the hydrophobic character of an otherwise highly polar side-chain. The introduction of the methyl groups leads to an increased burial of apolar ASA on complex formation, which results in more favorable solvation entropies. This interpretation is supported by the agreement between the measured differences in the entropic contributions to binding ($-T\Delta\Delta S$ = -1.9, -2.8 and -3.3 kcal·mol^{-1} for $H3_{10}$K4me1,- me2 and -me3, respectively, referenced to $H3_{10}$K4me0), and the differences in solvation entropy calculated (33) in terms of changes in accessible surface area ($-T\Delta\Delta S_{solv}$ = -1.7, -2.9 and -2.6 kcal·mol^{-1}).

3.4 ING4 reading of the histone code

The structural basis for the binding of ING4 to histone-3 N-terminal tail lies on numerous interactions that occur when the N-terminal histone residues occupy a depression on one side of the PHD. This results, however, in a low binding affinity which increases on methylation at K4. The relevance of the differences in the affinities for mono-, di- or trimethylated peptides is however unclear since peptide elongation at the C-terminus stabilizes the complexes blurring those differences. This is in contrast with ING2, which does discriminate between the different methylation states of $H3_{12}$K4meX, with

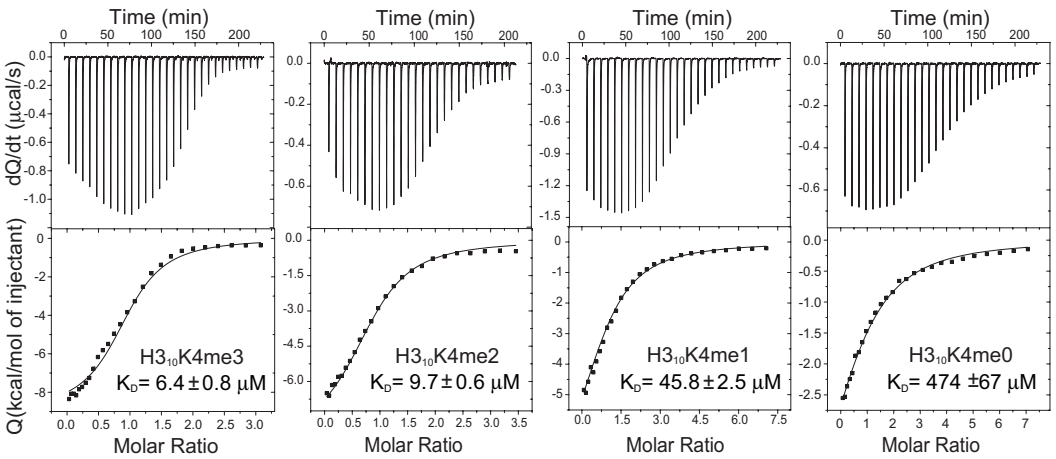

Figure 6. Calorimetric titrations of ING4-PHD with $H3_{10}$K4meX peptides. Shown are examples of the direct titrations of the $H3_{10}$K4me3, $H3_{10}$K4me2 and $H3_{10}$K4me1 peptides together with the displacement experiment for $H3_{10}$K4me0 using $H3_{10}$K4me1 as competing ligand. In all cases, upper panels represent the heat effect associated with the peptide injections and lower panels represent the ligand concentration dependence of the heat released upon binding, after normalization and correction for the heats of dilution. The symbols correspond to the experimental data and the continuous line to the best fit to a model of one set of identical binding sites (according to equation 1) for $H3_{10}$K4me3, 2 and 1, and to a complete displacement model for $H3_{10}$K4me0.

approximately ten-fold affinity increments for every additional methyl (14). The differences in the binding mode between the two ING proteins discussed above may explain these dissimilar affinities, which could be related with a different function. ING2, as part of the mSin3A-HDAC1 complex, links H3K4me3 recognition with transcriptional repression through histone deacetylation. ING4, involved in a complex containing the histone acetyltransferase HBO1 (11), may link the recognition of H3K4me3 (and of H3K4me2 and -me1) with transcriptional activation through histone acetylation.

The side chain of H3R2 lies in a groove that is separated by W221 from another groove occupied by H3K4me3 (Fig 1A), as was described for $H3_{15}$K4me3 bound to the double chromodomain of CHD1 (38). The affinity of the chromodomain binding to H3K4me3 was found to experience a four-fold reduction when the simultaneous asymmetric dimethylation at H3R2 (H3R2me2a) occurred on histone 3 tail. In ING4, the side chain of H3R2 interacts with the side chain of E220, while the corresponding position in CHD1 is occupied by a smaller glycine residue. Thus steric clashes may cause an even larger reduction in the affinity of ING4 for K4me3 when the simultaneous H3R2me2a modification occurs on the same histone site. The relevance of this discrimination lies in the mutual exclusion of H3K4me3 and H3R2me2a recently reported in actively transcribed genes (39,40). The measurement of the binding of ING4 to H4R2me2aK4me3 shows a six-fold reduction in binding affinity (Table 2 and supplemental Fig 4), indicating that ING4 can discriminate between chromatin regions enriched in one or both histone modifications. Dimethylation of H3R2 inhibits the recognition of H3K4me2 by the Spp1-PHD, a subunit of the Set1p histone methylation complex (40), thus regulating trimethylation at H3K4. Sequence homology suggests that the binding site for H3R2 in Spp1 is similar to ING4, indicating that dimethylation of H3R2 will also inhibit binding of ING4 to H3K4me3, which is consistent with the measured reduced affinity for the corresponding histone fragment. In this way, H3R2me2 may help to maintain the chromatin silent not only by inhibiting the recruitment of methyltransferases, but also of histone acetyltransferase complexes such as HBO1. Therefore, understanding how different combinations

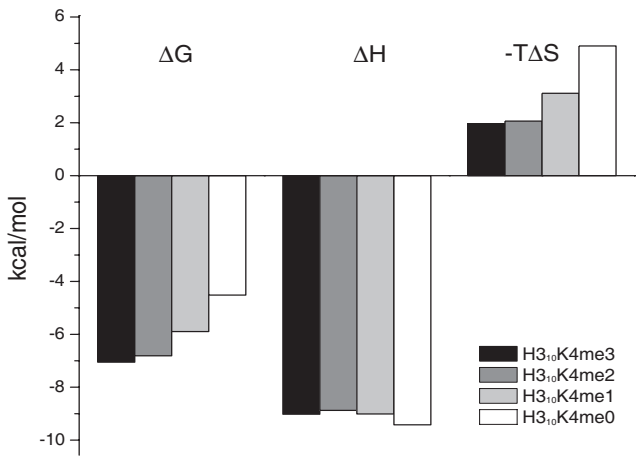

Figure 7. Thermodynamic parameters of the ING4-PHD/H3K4meX binding. The bar diagram shows the change in free energy (ΔG), enthalpy (ΔH), and in the contribution of entropy to the free energy (-TΔS) of the binding of $H3_{10}$K4meX peptides to ING4(188-249), as determined by ITC. The uncertainty in the experimental values is estimated to be about 5 %.

of histone modifications are read by their recognition modules may be critical to appreciating the regulatory mechanisms exerted through those modifications, as illustrated by the antagonizing role of H3K9me3 and H3S10ph in the recruitment of HP1 to discrete regions of the chromatin, thereby regulating gene expression (41,42).

The double chromodomain of CHD1 binds to H3K4me3 with similar affinities when combined with either H3K9me3, H3K9ac or H3S10ph, but with a twenty-fold reduced affinity when H3T3ph is present (38). Based on the structure of the ING4/H3K4me3 complex, phosphorylation of H3T3, a modification correlated with transcription activation, would generate a larger and negatively charged residue which could be accommodated by a conformational change at the side chain of K232, with which a favourable electrostatic interaction could be established (Fig 1A). H3K9me3 binds to ING4 with the same affinity as unmodified H3 (13), suggesting the compatibility of methylation at both lysine residues. Acetylation, however, removes the positive charge of H3K9 precluding its interaction with the aromatic side chain of Y206 (Fig 1D), and probably reducing the affinity of ING4 binding to H3K4me3K9ac. Phosphorylation of H3S10 would disrupt the hydrogen bond of its side chain with E195 (Fig 1D) and introduce an unfavourable electrostatic interaction, suggesting that it would be incompatible with H3K4me3 recognition. A similar effect, though reduced, might occur if H3T11 is the phosphorylated residue since its location on the PHD is close to the negatively charged N-terminal region of the PHD (Fig 3B and 4B). On the other hand, monomethylation in Arg8, which has been associated with repressed transcription, is probably of little relevance for H3K4me3 recognition since its side chain can adopt two different conformations (Fig 1) and the structure of the complex does not suggest additional interactions with H3R8me1. Our results suggest that H3K4me3 recognition by ING4 is compatible with H3T3ph, H3R8me1, and H3K9me3, while it is not compatible, or results partially impaired, by H3R2me2a, H3K9ac, H3S10ph, and H3T11ph. Further experimental studies on PHD fingers and other protein modules will provide more insights into the impact of multiple combinations of histone modifications and their role in the dynamic regulation of chromatin.

Acknowledgements

This work was supported by Fundación de Investigación Médica Mutua Madrileña grant to FJB, Ministerio de Educación y Ciencia (MEC) grant BFU-2005-02403 to GM, and MEC grant GEN2003-20642-C09-02 to GM and FJB. DPU and MJM were supported by a MEC Juan de la Cierva contract and by an EMBO long-term fellowship, respectively.

References

1. Li, B., Carey, M., and Workman, J. L. (2007) Cell 128, 707-719
2. Groth, A., Rocha, W., Verreault, A., and Almouzni, G. (2007) Cell 128, 721-733
3. Downs, J. A., Nussenzweig, M. C., and Nussenzweig, A. (2007) Nature 447, 951-958
4. Rountree, M. R., Bachman, K. E., Herman, J. G., and Baylin, S. B. (2001) Oncogene 20, 3156-3165
5. He, G. H., Helbing, C. C., Wagner, M. J., Sensen, C. W., and Riabowol, K. (2005) Mol Biol Evol 22, 104-116
6. Russell, M., Berardi, P., Gong, W., and Riabowol, K. (2006) Exp Cell Res 312, 951-961
7. Soliman, M. A., and Riabowol, K. (2007) Trends Biochem Sci 32, 509-519
8. Campos, E. I., Chin, M. Y., Kuo, W. H., and Li, G. (2004) Cell Mol Life Sci 61, 2597-2613.
9. Aasland, R., Gibson, T. J., and Stewart, A. F. (1995) Trends Biochem Sci 20, 56-59.
10. Bienz, M. (2006) Trends Biochem Sci 31, 35-40
11. Doyon, Y., Cayrou, C., Ullah, M., Landry, A. J., Cote, V., Selleck, W., Lane, W. S., Tan, S., Yang, X. J., and Cote, J. (2006) Mol Cell 21, 51-64
12. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. (2006) Nature 442, 100-103
13. Palacios, A., Garcia, P., Padro, D., Lopez-Hernandez, E., Martin, I., and Blanco, F. J. (2006) FEBS Lett 580, 6903-6908
14. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) Nature 419, 407-411
15. Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G., Shi, Y., Cote, J., Chua, K. F., and Gozani, O. (2006) Nature 442, 96-99
16. Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C., and Allis, C. D. (2006) Nature 442, 86-90
17. Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006) Nature
18. Otwinowski, Z., Minor, W., and Charles W. Carter, Jr. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. In. Methods in Enzymology, Academic Press
19. Read, R. J. (2001) Acta Crystallogr D Biol Crystallogr 57, 1373-1382
20. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr D Biol Crystallogr 53, 240-255
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr A 47 (Pt 2), 110-119
22. Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J., and Perrakis, A. (2004) Acta Crystallogr D Biol Crystallogr 60, 2222-2229
23. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Biochemistry 33, 5984-6003
24. Renner, C., Schleicher, M., Moroder, L., and Holak, T. A. (2002) J Biomol NMR 23, 23-33

25. Johnson, B. A. (2004) Methods Mol Biol 278, 313-352
26. Pawley, N. H., Wang, C., Koide, S., and Nicholson, L. K. (2001) J Biomol NMR 20, 149-165
27. Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995) J. Am. Chem. Soc. 117, 12562-12566
28. Bruschweiler, R., Liao, X., and Wright, P. E. (1995) Science 268, 886-889
29. Cole, R., and Loria, J. P. (2003) J Biomol NMR 26, 203-213
30. Palmer, A. G., Rance, M., and Wright, P. E. (1991) J. Am. Chem. Soc. 113, 4371-4380
31. Lipari, G., and Szabo, A. (1982) J. Am. Chem. Soc. 104, 4559-4570
32. Sigurskjold, B. W. (2000) Anal Biochem 277, 260-266
33. Luque, I., and Freire, E. (1998) Methods Enzymol 295, 100-127
34. Jacobs, S. A., and Khorasanizadeh, S. (2002) Science 295, 2080-2083
35. Pascual, J., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. (2000) J Mol Biol 304, 723-729.
36. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007) Nat Struct Mol Biol 14, 1025-1040
37. Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S., and Waters, M. L. (2007) Proc Natl Acad Sci U S A 104, 11184-11188
38. Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005) Nature 438, 1181-1185
39. Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlautz, H., Luscher, B., and Amati, B. (2007) Nature
40. Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bahler, J., Green, R. D., and Kouzarides, T. (2007) Nature
41. Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. (2005) Nature 438, 1116-1122
42. Hirota, T., Lipp, J. J., Toh, B. H., and Peters, J. M. (2005) Nature 438, 1176-1180
43. Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M., and Jones, T. A. (2001) Around O., Kluwer Academic Publishers, Dordrecht
44. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Journal of Applied Crystallography 26, 283-291
45. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002) Acta Crystallogr D Biol Crystallogr 58, 1955-1957

Supplementary material

Backbone Dynamics from ^{15}N Relaxation of Free and Bound ING4-PHD

The ^{15}N relaxation parameters measured for the free and complex forms of ING4(188-249) are represented in Figure 2. The heteronuclear $\{^1H\}$-^{15}N NOE, and the longitudinal (T_1) and transverse (T_2) relaxation times could be measured for 51 out of the 63 residues of free ING4-PHD (all except the N-terminal residue, the six prolines, and Val191, Asp192, Asn194, Cys212, Lys232 and Lys236 whose signals overlap). For the ING4-PHD bound to $H3_{10}K4me3$ or to $H3_{15}K4me3$, the relaxation data could only be measured for 49 common residues due to an increased overlap (Val191, Asp192, Asn194, Thr197, Ile219, Trp221, Gly228 and Lys236,) with respect to the spectrum of free ING4-PHD.

The residues at the N-terminus (188-195), Gly235 and at the C-terminus (246-249) show small NOE values (< 0.65) in the three forms of ING4-PHD, indicating flexibility on fast times scales (ps to ns). Global rotational diffusion correlation times of 6.7 ns, 7.4 ns and 7.5 ns were obtained for ING4-PHD free and bound to $H3_{10}K4me3$ and $H3_{15}K4me3$, respectively. These values are larger than expected for molecular species of the size of the PHD and the complexes. A possible explanation is protein aggregation, however, ultracentrifugation on the isolated PHD unequivocally indicates that the protein is monomeric and the linewidths of the protein signals do not change appreciably upon binding to the peptide, indicating that aggregation does not occur (data not shown). Most likely, the large number of flexible residues in the N- and C-terminal ends contributes to the increased overall correlation time as has been found in other cases (1,2). To confirm this possibility, we built a three dimensional model of the complex with extended chain ends added to the crystal structure with a random conformation. Hydrodynamic calculations using the program HYDRONMR on this model supports this interpretation of the overall correlation times (5.6 ns and 6.9 ns were obtained for free ING4-PHD and the complex respectively).

The principal components of the ING4-PHD inertia tensor have relative values of 1.00:0.80:0.45. The corresponding values for the ING4-PHD bound to $H3_{10}K4me3$, determined from the crystal structure, are 1.00:0.89:0.43, indicating that the shape of both species deviate from that of a sphere and approach a prolate ellipsoid. In agreement with this finding, the diffusion tensor that better explained the relaxation data was also anisotropic, with different values for the diffusion constants parallel and orthogonal to the long axis of the molecule ($D_{\parallel}/D_{\perp}$= 1.4 for ING4-PHD and 1.5 for ING4-PHD bound to H3K4me3). Therefore, the relaxation data were analyzed to obtain the internal dynamics parameters assuming an axially anisotropic diffusion model for the overall rotational tumbling.

The relaxation data were analyzed by using the model-free formalism to calculate the corresponding dynamics parameters for the amide ^1H-^{15}N pair of each residue. Most of the ^{15}N relaxation parameters could be satisfactorily fitted to one of the two simplest modelswhich describe the internal dynamics of the molecule in terms of a generalized order parameter S^2 and an effective correlation time τ_e for fast motions. In a few cases (four residues in free ING4-PHD and two in the complex with to $H3_{10}K4me3$) it was necessary to include a contribution of slow motions to the transverse relaxation time, on the μs to ms time scale, which are characterized in terms of the conformational exchange rate R_{ex}, while for others (three residues in free ING4-PHD and four in the complex) the inclusion of the amplitude of the internal motions (S_f^2) was also necessary to obtain a good fitting. The values of S^2 is represented for each residue of free and bound ING4-PHD in Figure 5.

Isothermal titration calorimetry.

The binding isotherms were analyzed by non-linear least-square fittings of the experimental data to a model corresponding to a single set of identical sites, according to the following equation

$$Q = \frac{n \cdot [M_t] \cdot \Delta H \cdot V_0}{2} \cdot \left[1 + \frac{[X_t]}{n \cdot [M_t]} + \frac{1}{n \cdot K \cdot [M_t]} - \sqrt{\left(1 + \frac{[X_t]}{n \cdot [M_t]} + \frac{1}{n \cdot K \cdot [M_t]} \right)^2 - \frac{4 \cdot [X_t]}{n \cdot [M_t]}} \right]$$

where Q is the net heat of binding, n is the number of binding sites, ΔH is the change in enthalpy of the binding reaction, K is the association constant, V_0 is the active cell volume, and $[M_t]$ and $[X_t]$ are the total concentrations of macromolecule and ligand respectively.

Figures

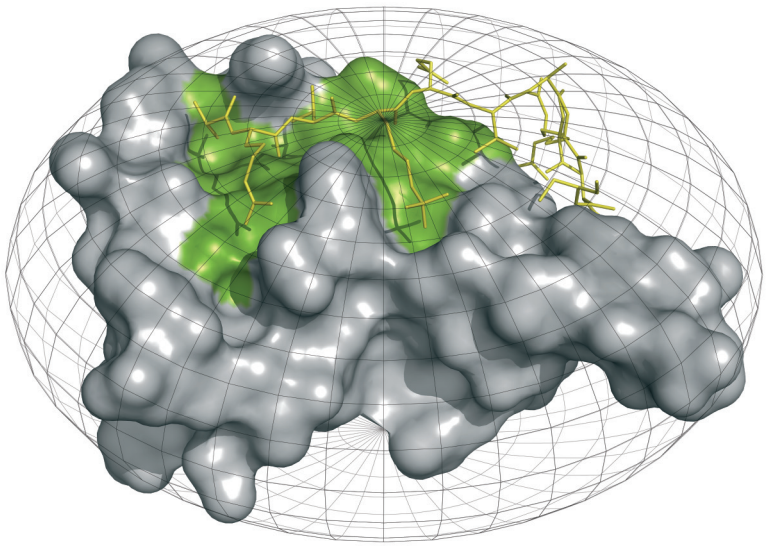

Figure 1. Surface representation of ING4-PHD bound to H3K4me3 peptide (in yellow sticks) inserted in an oblate ellipsoid. The amino acids of ING4 that form the three grooves in the surface of PHD are coloured in green.

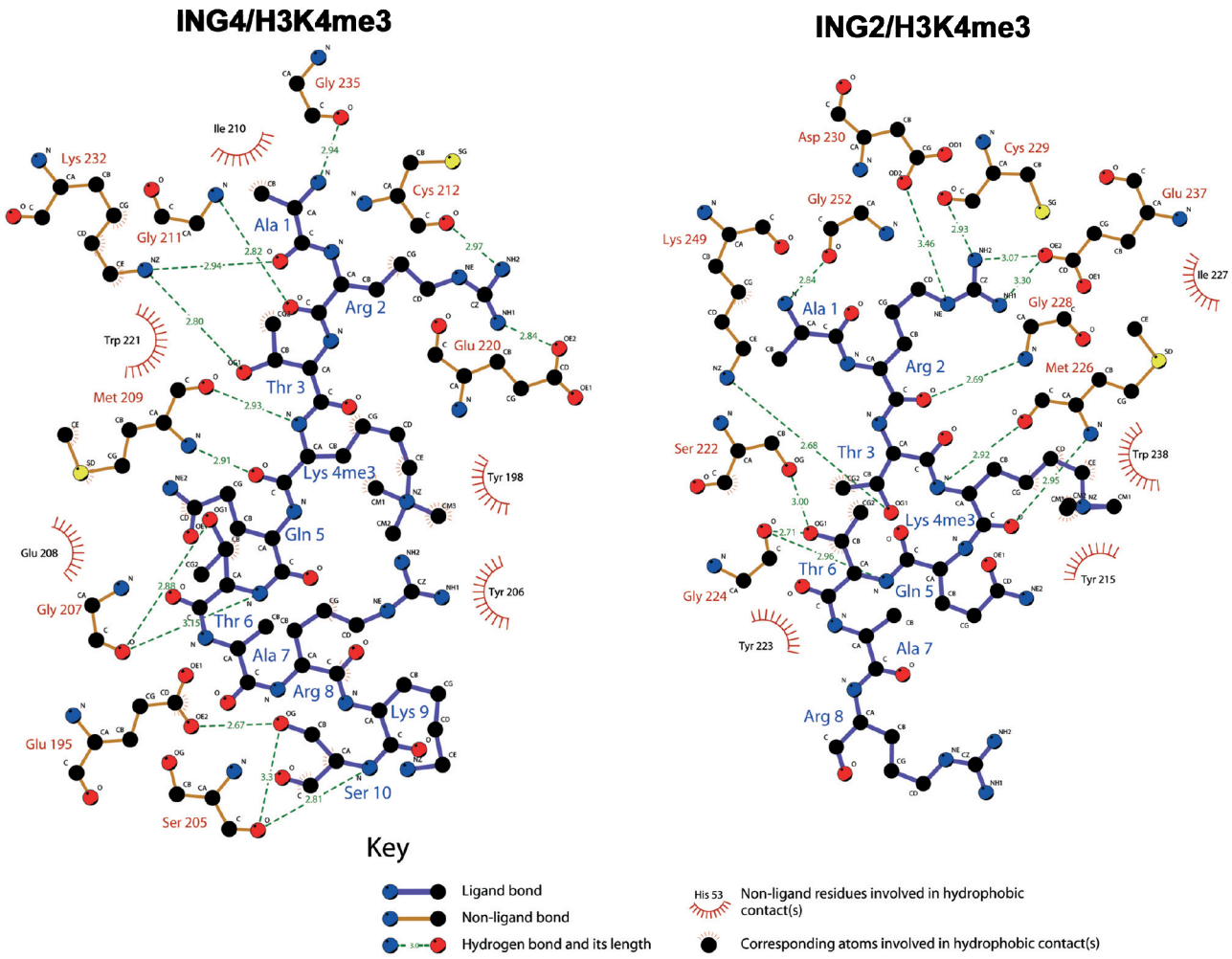

Figure 2. Schemes of the interactions of $H3_{10}K4me3$ peptides with the PHD of ING4 and ING2. The figures were prepared with Ligplot (3) and Pymol (http://pymol.sourceforge.net) using PDB entries 2VNF for ING4 and 2G6Q for ING2.

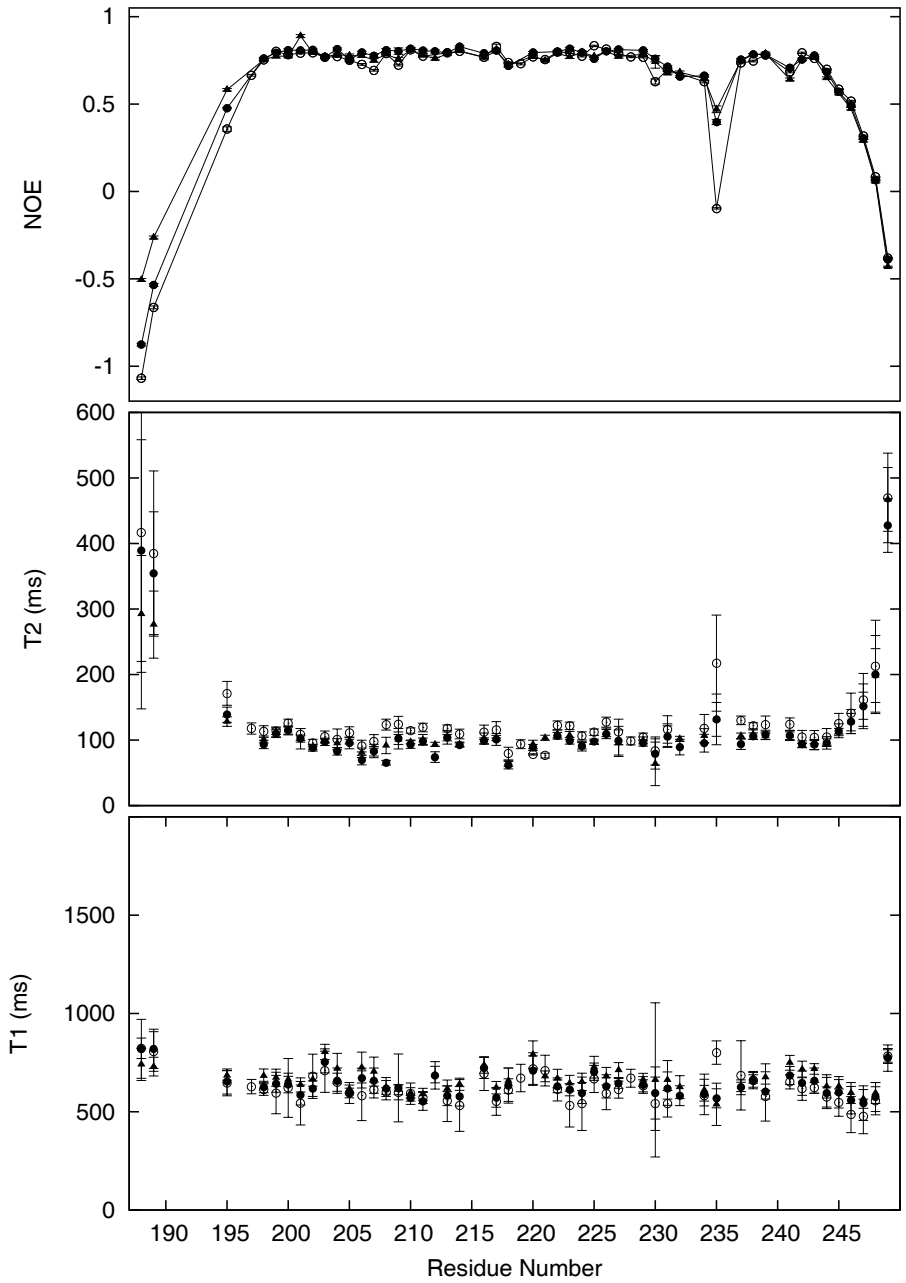

Figure 3. Backbone NMR relaxation data for ING4(188-249) free and bound to H3K4me3 peptide. Heteronuclear {^{1}H}-^{15}N NOEs, and ^{15}N transversal (T_2) and longitudinal (T_1) relaxation times are represented for each residue of PHD in its free form (open circles) and in complex with $H3_{10}$K4me3 (closed circles) or $H3_{15}$K4me3 (closed triangles).

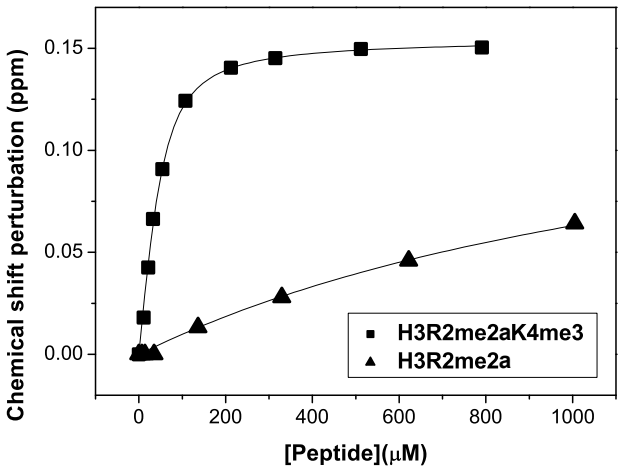

Figure 4. NMR analysis of the binding of ING4-PHD to $H3_{10}$R2me2aK4me3 and $H3_{10}$R2me2a peptides. The CSP of the C212 amide resonance of ING4(188-249) in ^{1}H-^{15}N-HSQC spectra is represented as a function of peptide concentration. The continuous line is the fitting to the binding model (4). The height of the symbol indicates the experimental error. In these curves the CSP of C212, instead of W237 are represented and fitted. The reason for this choice was that in order to fit the data for peptide $H3_{10}$R2me2a, for which the maximum CSP is not easily accessible experimentally, it is necessary to assume that it has the same value as measured for $H3_{10}$R2me2aK4me3 peptide, for which saturation is achieved experimentally. Therefore, the use of a residue far away from H3K4 (methylated in the former peptide but not in the later one) was needed. C212 was sufficiently distant from H3K4 but sufficiently perturbed by peptide binding to serve this purpose. The dissociation constant measured for $H3_{10}$R2me2aK4me3 peptide using W237 or C212 was the same within experimental error (19.2 ± 1.7 μM and 17 ± 2 μM respectively).

References

1. Gonzalez, C., Neira, J. L., Ventura, S., Bronsoms, S., Rico, M., and Aviles, F. X. (2003) Proteins 50, 410-422
2. Viles, J. H., Donne, D., Kroon, G., Prusiner, S. B., Cohen, F. E., Dyson, H. J., and Wright, P. E. (2001) Biochemistry 40, 2743-2753
3. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) Protein Eng 8, 127-134
4. Palacios, A., Garcia, P., Padro, D., Lopez-Hernandez, E., Martin, I., and Blanco, F. J. (2006) FEBS Lett 580, 6903-6908

THE TUMOUR SUPPRESSOR ING4 IS A BIVALENT RECOGNITION MODULE OF H3K4me3 THAT DIMERIZES THROUGH ITS N-TERMINAL DOMAIN WITH A COILED STRUCTURE

Alicia Palacios[$], Alberto Moreno[†], Bruno Oliveira[‡], Teresa Rivera[‡], Jesús Prieto[‡], Pascal García[‡], Rosario Fernández[¶], Pau Bernadó[%], Ignacio Palmero[†] and Francisco J. Blanco[$]

[$]*Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia, 48160 Derio, Spain;* [‡]*Structural Biology and Biocomputing Programme, Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain, Melchor Fernández. Almagro 3, 28029-Madrid, Spain;* [¶]*Centro de Biología Molecular Severo Ochoa, UAM-CSIC, Madrid;* [†]*Instituto de Investigaciones Biomédicas "Alberto Sols", CSIC-UAM, Arturo Duperier 4, 28029 Madrid, Spain; and* [%]*Institute for Research in Biomedicine, Baldiri Reixac, 10-12, 08028-Barcelona, Spain.*

Abstract

The inhibitor of growth 4 (ING4) is a 29 kDa nuclear protein of the ING family of tumour suppressors. Sequence homology and phylogenetics indicate that ING proteins have a similar architecture, with conserved N- and C-terminal regions, and a central non conserved region with the nuclear localization signal (NLS). The C-terminal plant homeodomain (PHD) is the site for recognition of histone H3 methylated at residue Lys4, and has a basic tail of different lengths that, in ING1 and ING2 but not ING4, acts as a receptor for phosphoinositides. To experimentally characterise the domain structure and organization of ING4, several constructs and the full length protein have been produced and analysed. We show that ING4 forms homodimers *in vitro*, as a pure protein in solution and also in living cells. The dimerization site is at the N-terminal domain of ING4, with a helical coiled coil structure. A direct interaction of the NLS region with p53, reported previously based on pull-down experiments, is not observed by NMR analysis of mixtures of the pure proteins. Full length ING4 binds trimethylated histone H3 tail at the same site and with the same affinity as its isolated PHD finger, making it a bivalent reader of the H3K4me3 histone modification.

Key words: ING4, inhibitor of growth, domain structure, PHD, methylated histones, tumour suppressor.

Abbreviations: ING, inhibitor of growth; NMR, nuclear magnetic resonance; HSQC, heteronuclear single quantum coherence spectroscopy; CD, circular dichroism; SAXS, small-angle X-ray scattering D1, fragment 1-118; D2 fragment 119-187; D3 fragment 188-249; D1D2, fragment 1-188; D2D3, fragment 119-249; NLS, nuclear localization signal; PHD, plant homeodomain; LZL, Leucine Zipper-Like motif.

1. Introduction

The ING family of proteins restricts cell growth and proliferation, induce apoptosis, participate in DNA repair, and cell migration, modulate cell cycle progression and down-regulate the transcription of certain genes that are markers of different types of cancer (1-4). Since the identification of ING1 in 1996 (5), four human homologues have been found and named ING2-5 (6). Their involvement in key cellular processes is manifested in the presence of orthologs of the human proteins in mouse, rat, fish, mosquito, fruit fly, worm, fungi and plants (6). Their mechanism of action is multiple and varies from protein to protein, although with common features amongst them. They interact with several nuclear proteins, and form part of stable histone acetylation and deacetylation complexes (7), which accounts for the regulation of gene expression by ING through the modulation of chromatin structure. ING1 and ING2 also act as nuclear phosphoinositide receptors and could be involved in the regulation of nuclear responses to DNA damage (8).

The amino acid sequences of ING proteins suggest that they are organized into three regions: a conserved N-terminal region, a central non-conserved region, with a Nuclear Localization Signal (NLS) and a highly conserved C-terminal region with homology to the structural domain *Plant HomeoDomain* (PHD). A detailed phylogenetic analysis has identified several sequence motifs in the three regions (6). According to these motifs, the five ING proteins can be classified in two groups. On one hand ING1

and ING2, whose NLS has three nucleolar translocation motifs, and which have a positive charge rich sequence at the end of the PHD. On the other hand, the group of ING3-5, which do not show a conservation of these motives. This clustering is consistent with the phylogenetic tree of the family (6) and with the observation that human ING1 and ING2 copurifiy with histone deacetylase complexes, while ING3,4 and 5 copurify with histone acetyl transferase complexes (7). ING2 and the second group share a sequence pattern at the N-terminus that has been termed Leucine Zipper-Like motif (LZL), with leucine residues separated by approximately seven residues, and that could be a site for homo- or hetero-oligomerization with other proteins containing similar motifs (6). Also at the N-terminus there is a conserved sequence (potential chromatin regulatory, or PCR) that has been hypothesized to be a site of binding to HAT/HDAC complexes.

The three-dimensional structure of ING proteins is not known, but the structure of the conserved C-terminal regions homologous to PHD has been determined for ING1(9), ING2 (10, 11), ING4 (12, 13) and ING5 (14). They adopt the folding characteristic of this domain, of about 60 amino acids, with a double Zn^{2+} ion binding motif of the type C4HC3 and a globular structure but little regular secondary structure (15). The PHD has been detected in more than 400 eukaryotic proteins, many of them involved in gene expression regulation and chromatin remodelling (16). The PHD fingers of the five ING proteins recognise histone H3 tails methylated at Lys4 (H3K4me3), linking this posttranslational histone modification to chromatin remodelling and transcription regulation (9-11, 14).

ING4 was identified in a search for homologues to ING1, 2 and 3 (17). Its function has not been as extensively studied as ING1, the best-known member for the family, however certain singular aspects of its function, such as its links with migration and angiogenesis, have been characterised (18-20). ING4 regulates the growth of gliomas, where a reduction in its expression levels is observed and correlates with the progression from lower to higher grades of tumors (18). The activation of ING4 in glioblastoma cells transplanted into mice reduced tumor growth and vascularization, while its suppression accelerated growth and generated tumors with a robust vasculature. Similar anti-angiogenic or anti-invasive effects of ING4 have recently been reported for other tumor types, such as melanoma or myeloma (21, 22). This anti-angiogenic activity of ING4 seems to be unique in the family and occurs through the transcriptional repression of proangiogenic genes regulated by the NFκB transcription factor (18, 22). ING4 was also detected in a search for genes that suppressed the loss of contact inhibition, a characteristic *in vitro* property of cancerous cells (23). This and other studies suggested that ING4 suppresses tumour biology without directly deregulating the cell cycle (24), although other reports suggest that it induces cell cycle arrest (17, 25). The ING4 locus is deleted in 10-20% of primary tumours and breast cancer cell lines (23) and at least 10 mutations have been detected in tumour derived cell lines (23). Seven out of the ten are located in the NLS, possibly interfering with the nuclear transport of ING4 and with its interaction with p53 (26). Recently we have reported the molecular basis of the H3K4me3 recognition by the PHD finger of ING4 (13). Methylated versus non-methylated histone tails discrimination by ING4 is predominantly of entropic nature likely due to solvation effects, but there is little differentiation among the three possible methylated states. Here we experimentally characterise the overall structural organization of the full-length ING4 molecule, and show that it forms homodimers through its N-terminal coiled coil domain and recognizes histone methylated peptides through the same binding site and with the same affinity as its isolated PHD finger, making it a bivalent reader of H3K4me3 mark.

2. Materials and Methods

2.1 Sequence analysis and structure prediction

The sequence alignment of the five ING protein sequences (as obtained from the Swissprot database) was performed with A Clustal X (27). For ING1 the predominant isoform in healthy tissues (p33ING1b) was used (28). Secondary structure prediction was performed using JPred (29). Disordered regions were predicted with PONDR (30).

2.2 Cloning and mutagenesis

A synthetic gene of ING4 (isoform v1, (19)) with codons optimized for expression in *E. coli* (Entelechon GmbH) was used to subclone five different ING4 constructs into the expression vector pET11d by PCR using the appropriate forward and backward oligonucleotides. The full length ING4 and four fragments designed to encompass one or two of its three regions or domains were cloned: D1D2D3 (full-length ING4, residues 1-249), D1 (1-118), D1D2 (1-188), D2D3 (119-249), and D3 (189-249). For overexpression of ING4 in 293T cells, the ING4 human gene was amplified by PCR with N-terminal AU5 or HA tags, and cloned into the retroviral expression vector pLPC.

2.3 Protein expression and purification

All the proteins were over-expressed in *E. coli* BL21(DE3) cells grown in LB medium supplemented before induction with 50 μM $ZnCl_2$. Isotope enriched proteins were produced in minimal medium with $^{15}NH_4Cl$. Expression was induced with 0.5 mM IPTG at an OD_{600} of 0.6 and the cells were harvested by centrifugation after 4 hours of induction at 37 ºC. After sonication and ultracentrifugation, proteins were found in the soluble (D1), insoluble fraction (ING4, D2D3 and D3) or in both (D1D2). Insoluble proteins were recovered from the pellets by solubilisation in 6 M urea, refolded by a 1:10 dilution into cold 20 mM, Tris pH 8.0, 1 mM DTT and 50 μM $ZnCl_2$ (for constructs containing the PHD) and purified by anion-exchange chromatography (Q-Sepharose column, GE Healthcare), followed by gel filtration separation (Superdex 75, GE Healthcare). The purification of insoluble protein D2D3, was facilitated by decreasing the pH of the soluble fraction in 6 M urea to 3.0 to precipitate acid proteins and DNA. After ultracentrifugation the pH was changed back to 8.0 and purified by anion exchange (the protein comes in the flowthrough) followed by cation exchange chromatography (SP-Sepharose column, using a NaCl gradient for protein elution). After refolding by dilution into cold buffer the protein was concentrated and polished by gel filtration. Soluble D1 and D1D2 proteins were purified by anion-exchange chromatography followed by several steps of gel filtration separation. A small amount of D3 was purified directly from the soluble fraction of the lysed cells by anion exchange and gel filtration chromatography yielding an identical 1D NMR spectrum as the refolded molecule. Pure proteins were concentrated and dialyzed against 20 mM phosphate buffer pH 6.5, 50 mM NaCl and its identity verified by mass spectrometry.

The preparation of labelled and unlabelled p53 was done as described (31).

2.4 Analytical ultracentrifugation

Ultracentrifugation experiments were performed at 20 °C in an Optima XL-A (Beckman-Coulter Inc.) analytical ultracentrifuge equipped with UV-visible optics, using an An50Ti rotor, with 3 mm double sector centerpieces of Epon charcoal. The sedimentation velocity experiment was carried out at 42,000 rpm, and absorbance scans were taken at 280 nm. These measurements were performed in: 20mM sodium phosphate buffer pH 6.5, 200mM NaCl, 1mM DTT (D1, D1D2, D2D3); 20mM Tris pH 8.0, 300mM NaCl, 1mM DTT (D1D2D3); 20mM sodium phosphate buffer pH 6.5, 50mM NaCl, 1mM DTT (D3); phosphate buffer saline pH 7.2, 300mM NaCl, 1mM DTT (D1-M50V and D1). Concentrations used for each sample had OD values between 0.5 and 0.7 at 280 nm. The sedimentation coefficients were calculated by continuous distribution c(s) Lamm equation model as implemented in the SEDFIT program. These experimental sedimentation values were corrected to standard conditions to get the corresponding $s_{20,w}$ values using the SEDNTERP program (32). Further hydrodynamic analysis (i.e. calculation of frictional coefficient ratio) was performed with the SEDFIT program to obtain de c(M) distribution (33).

Equilibrium measurement of ING4 full-length was performed at 0.35 OD values at 280 nm in the same buffer as the sedimentation velocity measurement. Short column (23 µl), low speed sedimentation equilibrium was performed at three successive speeds (5,000, 7,000 and 11,000 rpm), and the equilibrium scans were taken (after 20 h) at a wavelength of 280 nm. The system was assumed to be at equilibrium when successive absorbance scans did not change. The base-line signal was measured after high speed centrifugation (5 h at 42,000 rpm). The apparent average molecular weight of the protein was obtained using the program EQASSOC (34).

2.5 Chemical crosslinking

Glutaraldehyde (GTA) at 25% was purchased from Sigma. Experiments were performed at room temperature with D1, and D3 at concentrations of 25 µM in phosphate buffer saline (PBS, 10 mM sodium phosphate, 2 mM potassium phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride, pH 7.2) plus 1mM DTT and supplemented with NaCl to a total concentration of 300 mM. Proteins were incubated with three different protein:GTA molar ratios (1:20, 1:50 and 1:80) for 1 hour and reactions were stopped with 2 M $NaBH_4$ for 15 min. Crosslinked proteins were visualized in SDS-PAGE.

2.6 Circular dichroism (CD)

CD spectra were performed in a Jasco J-810 spectropolarimeter at 25 ºC. The spectra were the average of 15 scans, recorded using a 0.1mm path length quartz cuvette on samples prepared by dialysis against 20mM sodium phosphate buffer pH 6.5, 200 mM NaCl, 1mM DTT at protein concentrations of 50 µM. Thermal denaturation curves were measured in 2 mm path length cuvettes closed with a teflon cap on 25 µM protein solutions and the CD signal measured at 222 nm. Equilibrium unfolding was induced increasing temperature at a rate of 1 °C/min in the range 5 ºC to 95 ºC and the ellipticity at 222 nm was recorded at intervals of 1 °C. The average helical content calculated from the molar residue ellipticity at 222 nm assuming 50 residue long helices (as predicted by J-Pred) is 30, 41 and 63 % for ING4, D1D2 and D1, respectively (35).

2.7 NMR spectroscopy

NMR experiments were recorded on Bruker AVANCE 600, 700 and 800 with triple resonance cryoprobes at 298 K. NMR samples were prepared with a protein concentration of approximately 130 μM in 20 mM sodium phosphate pH 6.5, 200 mM NaCl 1mM perdeuterated dithiothreitol (d_{10}-DTT), 5% $^{2}H_2O$ and 0.01% NaN_3. Chemical shifts were measured relative to internal DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) for ^{1}H and calculated for ^{15}N (36). Spectra were processed with XWINMR, TOPSpin (Bruker) and/or NMRPipe (37).

2.8 Ligand binding analysis by NMR

The synthetic peptides H3K4meX, correspond to residues 1-15 of histone H3 plus an extra tyrosine residue at the C-terminus (ARTKQTARKSTGGKAY) to measure peptide concentration by ultra violet (UV) absorbance. Three peptides named with X= 1, 2 or 3 (mono-, di-, and trimethylated, respectively) differ in the number of methyls bound to the amino group of Lys4. These peptides have free N- and C-termini and were purchased from NeoMPS. The titration of ING4 with the peptides and the NMR binding analysis was performed as described (12).

^{1}H-^{15}N HSQC NMR experiments on ING4/p53 samples were measured at molar ratios of 1:1 at 20 ºC. For p53 observation U-^{15}N-^{2}H full-length p53 (with protonated amide groups) was used and TROSY-HSQC spectra recorded at 900 MHZ. For D2D3 observation U-^{15}N D2D3 samples HSQC spectra were recorded at 600 MHZ. Three different experiments were designed in order to explore the interaction of ING4 with p53: labelled-p53/ING4, labelled-p53/D2D3 and labelled-D2D3/unlabelled p53. The control samples (the isotopically labelled proteins alone) and the test samples (mixtures) were simultaneously dialyzed against the same buffer (25 mM sodium phosphate pH 7.2, 150 mM NaCl, 5 mM perdeuterated DTT) and the NMR spectra were acquired under the same conditions. The protein concentration was 100 μM in the labelled p53 containing samples and 50 μM in the labelled D2D3 containing sample (monomer protein concentration).

2.9 Immunoprecipitation.

Immunoprecipitation was carried out essentially as previously described (38). 293T cells were transiently transfected with vectors expressing AU5-tagged and HA-tagged versions of full-length human ING4, using a standard calcium phosphate protocol. Total protein lysates were prepared 48 hours after transfection with medium-salt lysis buffer as previously described (39). For immunoprecipitation, lysates containing 1mg of total protein were incubated with 1μl of an antibody against the HA tag (clone 12CA2, Roche) or AU5 tag (MMS135R, Covance) or without antibody as a negative control, overnight at 4 ºC, with constant rotation. Thirty microlitres of a slurry containing Protein A-agarose beads (GE healthcare) was added to the mix and incubated for 1h in the same conditions. Beads were washed four times with ice-cold lysis buffer and proteins in the immunoprecipitate were recovered by incubation with 2xSDS loading buffer (40μl per immunoprecipitation) for 5 min at 90 ºC. The presence of each protein in the immunocomplex was analysed by Western Blotting with antibodies against AU5-tag (1/500) and HA-tag (1/500).

2.10 Small angle X-ray Scattering

Small-angle X-ray scattering (SAXS) data were collected at X33 beamline at the European Molecular Biology Laboratory (EMBL) in the storage ring DORIS III of the Deutsches Elektronen Synchrotron (Hamburg) (40). Scattering curves of D1 and D1D2 contructions were measured at 10ºC at protein concentrations of 500 and 260 μM for D1, and 251 and 186 μM for D1D2. An exposure time of two minutes was used for measurements. The scattering profiles covered a range of momentum transfer of $0.02< s <0.5$ Å^{-1}. Radiation damage was monitored by repetitive two-minute exposures of protein solutions, and no significant changes were observed. Buffer scattering profiles were measured before and after the samples, they were averaged, and used for the subtraction from the protein scattering profiles. Resulting profiles were merged to avoid interparticle interactions. All data manipulations were done using standard procedures with the software PRIMUS (41). The forward scattering, I(0), and the radius of gyration, R_g, were evaluated using the Guinier approximation, assuming that at very small angles ($s<1.3/R_g$), the intensity can be well represented as $I(s)=I(0)\ \exp(-(sR_g)^2/3)$. The distance distribution function, *p(r)*, was computed from the entire curve with the program GNOM (42) by optimizing the maximum particle dimension, D_{max}. The apparent molecular weights of the proteins were estimated from their forward scattering, I(0), by comparison to the one obtained for a Bovine Serum Albumin sample of 5.8 mg/ml.

Low resolution *ab initio* reconstructions were built from D1 data using the program GASBOR (43). The protein was represented by an assembly of 236 dummy residues inside a search volume. Starting from a random conformation an assembly was used a simulated annealing protocol to fit the experimental scattering profile. Ten independent GASBOR reconstructions were performed, and the independent models were averaged with the program package DAMAVER (44) to yield the most probable shape.

3. Results and discussion

3.1 Structure prediction and construct design

To explore the domain organization of ING4, several constructs were defined taking into consideration the preservation of conserved sequence stretches, the avoidance of predicted secondary structure element disruption, and the particular residues found at the borders of the putative domains that may facilitate solubility and analysis of the expressed proteins. The summary of these analysis and defined constructs is shown in figure 1. JPred predicts the presence of two long helices (residues 4-52 and 57-107) in the N-terminal conserved domain, no secondary structure in the non-conserved central region and three short β-strands in the PHD finger, as seen in the solved structures of this PHD (12). The helices partially overlap

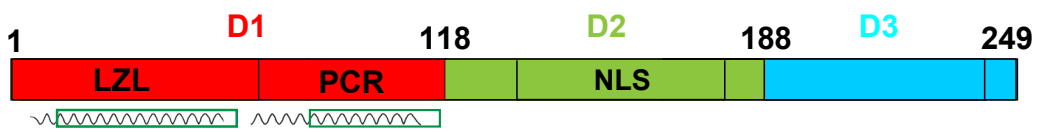

Figure 1. Scheme of the domain organization of ING4 with the boundaries of the constructs designed for their study. The predicted helices and coiled coil structures in the N-terminal domain are indicated with curvy black lines and dark green boxes, respectively.

with two segments (21-55 and 90-117) predicted to form coiled-coil structures by the Lupas algorithm (45), and with the postulated leucine zipper-like (LZL) motif (1-65) and putative oligomerization site (6). The boundaries of the N-terminal domain (D1) were thus defined as residues 1-118, which are followed by a long stretch of predominantly polar amino acids with high potential to be disordered, as predicted by PONDER. The C-terminal construct (residues 188-249, named hereafter D3) has been described previously (12), and consists of the canonical PHD, with its characteristic C4HC3 zinc finger motif, plus short chain extensions at the N- and C-termini which are not strictly part of the PHD finger and have been found to be flexible and disordered (13). The central region, rich in positively charged amino acids

Table 1.

Molecular weights obtained by sedimentation velocity experiments and SAXS to ING4 full-length and ING4 fragments, and RG and Dmax to D1 and D1D2 constructs observed by SAXS.

		Sedimentation velocity	Small angle X-ray scattering			
	Theoretical MW	Observed MW	Apparent MW	RG (Guinier)	RG (GNOM)	Dmax
ING4	28,530Da	55,500 Da	n.d.[a]	n.d.[a]	n.d.[a]	n.d.[a]
D1	13,857 Da	27,400 Da	30,000	26.4 ± 0.3	27.8	104
D1D2	21,479 Da	39,500 Da	90,000	41.5 ± 0.3	44.9	168
D2D3	14,692 Da	14,600 Da	n.d.[a]	n.d.[a]	n.d.[a]	n.d.[a]
D3	7,315 Da	6,800 Da	n.d.[a]	n.d.[a]	n.d.[a]	n.d.[a]

[a]Not determined

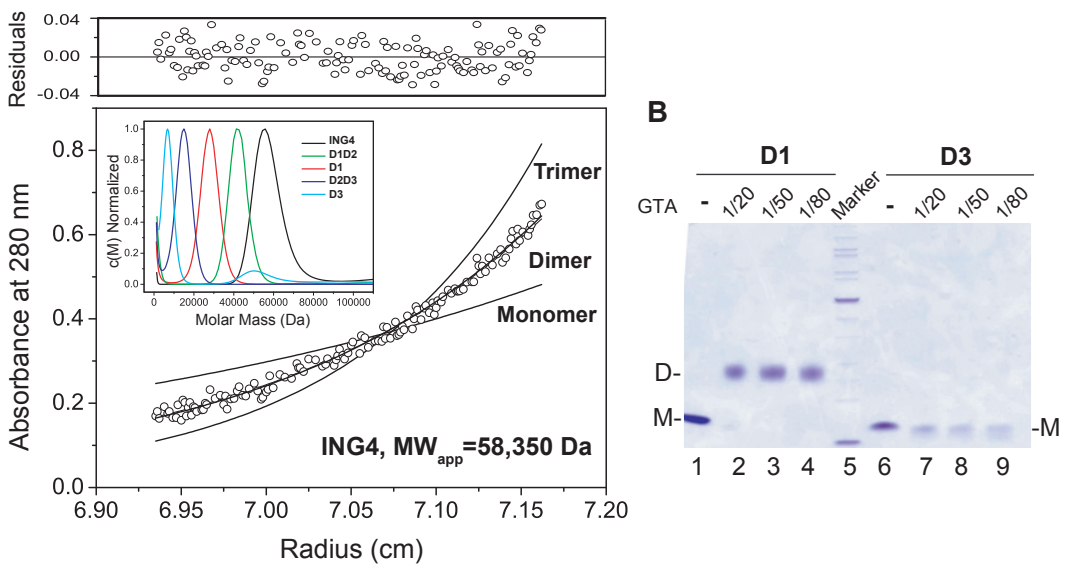

Figure 2. Oligomerization state characterization of ING4. A. Analytical ultracentrifugation measurements. Sedimentation equilibrium gradient of ING4 full-length at 11,000 rpm and 20 ºC. The open circles represent the experimental data and the solid lines represent the theoretical gradients of an ING4 full-length monomer, dimer, and trimer. The upper panel shows the residuals of the fitting of the equilibrium gradient to the sedimentation model for the dimeric species. Inset: sedimentation velocity distribution of the different constructs of ING4 at 42,000 rpm and 20 ºC. B. PAGE-SDS gel with coomassie blue staining of crosslinked product of constructs D1 and D3 of ING4. Lanes are (1) D1 control with no GTA added, (2) D1 with 1/20 protein:GTA ratio, (3) D1 with 1/50 protein:GTA ratio, (4) D1 with 1/80 protein:GTA ratio, (5) molecular weight markers, (6) D3 control with no GTA added, (7) D3 with 1/20 protein:GTA ratio, (8) D3 with 1/50 protein:GTA ratio, (9) D3 with 1/80 protein:GTA ratio.

contains the NLS (6) and is named hereafter D2 (residues 119-188), although no such isolated construct was studied. Its conformational behaviour was explored as part of the other constructs containing it.

3.2 ING4 dimerizes in vitro and in vivo.

The sedimentation equilibrium on ING4 (Figure 2) shows a mass distribution that cvan be fitted to a protein of and apparent molecular mass of 58,350 Da, a value very close to double the molecular weight of ING4 calculated from its amino acid sequence (57,060 Da). This result indicates that the protein is a dimer in solution. Sedimentation velocity measurements on ING4 and on its D1, D1D2, D2D3 and D3 fragments show that those constructs containing D1 are dimeric while those lacking D1 are monomeric (Table 1, figure 2A). The analysis of the forward scattering, I(0), measured by SAXS indicates that the apparent molecular weight for D1 and D1D2 correspond to dimeric species, see Table 1. These result are confirmed by cross-linking experiments of D1 and D3 (figure 2B), and demonstrate that D1 contains the dimerization site of ING4.

Sedimentation velocity on ING4 shows a distribution profile corresponding to a sedimentation coefficient of 2.9 S (data not shown), while for a 57 kDa globular protein, it a sedimentation coefficient of 5.1 S would be expected. The much smaller value observed indicates that the shape of the ING4 molecule differs substantially from a sphere. When modelled as an oblate ellipsoid, an axial ratio of 8 is required to give the observed sedimentation coefficient. These results are consistent with the elution volume of ING4 from calibrated gel filtration columns, which corresponds to an apparent molecular weight larger than expected for a globular protein of 57 kDa (data not shown).

Once we observed dimerization of purified ING4 proteins in solution, we wished to know the physiological relevance of this finding. To this end, we investigated whether ING4 has the abilty to form homodimers in live cells. 293T cells were transiently transfected with two vectors that direct the

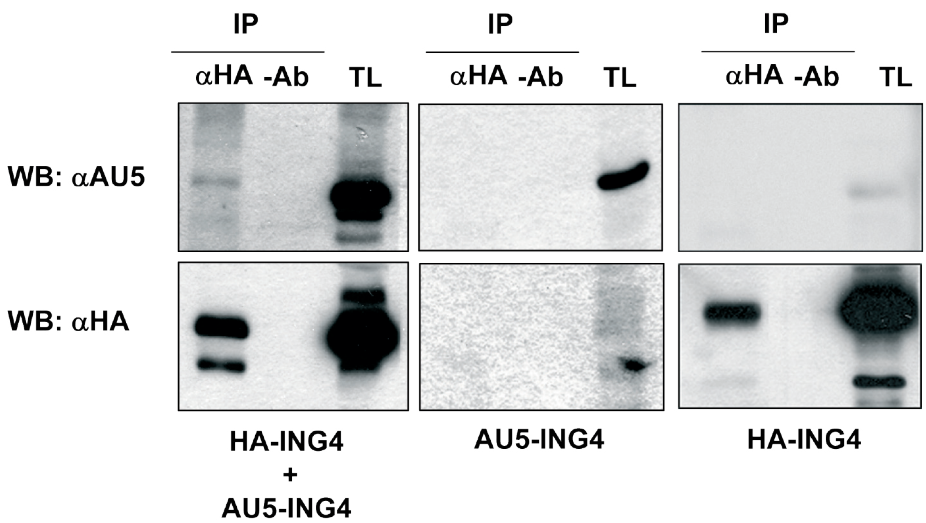

Figure 3. Immunoprecipitation showing dimerization in living cells. Lysates from 293T cells transiently transfected with vectors expressing HA-tagged and AUG-tagged ING4 (left panel), or each vector alone (middle and right panels) were immunoprecipitated with an antibody against the HA tag (αHA) or without antibody (-Ab) and the presence of each protein in the complexes was analysed by Western Blot (WB) with antibodies against each of the tags. TL is the total lysate.

expression of full length ING4 with two different tags, AU5 and HA. To test their ability to form dimers, total lysates from cells expressing both constructs, together or independently, were immunoprecipitated with antibodies against each tag, and the presence of the other construct in the complex was assayed by Western Blot (Figure 3). We found that the AU5-ING4 construct can be specifically co-precipitated with an antibody against the HA tag, only in cells with co-expression of the HA-ING4 construct (Figure 3, left panel), indicating the ability of the ING4 constructs to associate in vivo. These results confirm that the dimerization of ING4 observed with the purified proteins also occurs inside living cells (figure 3).

3.3 ING4 consists of two folded and independent domains linked by a flexible segment.

The far-UV circular dichroism spectra of the different ING4 constructs are shown in figure 4. The spectrum of D1D2D3 shows the typical signature of a high content of helical secondary structure, with minima at 208 and 222 nm, although the absolute ellipticity indicates that a large part of the chain is not helical (see below). The helical signature is more pronounced in the D1D2 construct and even more in the D1 fragment, whose molar ellipticity at 222 nm is larger (in absolute value) than at 208 nm ($[\Theta]_{222}/\ [\Theta]_{208} = 1.1$). This relative intensity is typical of coiled coil structures (46), and is consistent with the structure predictions described above. The spectrum of D3 has a minimum at 202 nm and a small negative ellipticity in the range 230-210, as expected from the low content of canonical α-helix and β-sheet secondary structure in the PHD finger, which consists mainly of loops stabilised by the two Zn^{2+} clusters (12). The shape of the spectra of D1D2 and D2D3 can be explained by the combination of the spectra of D1 or D3, respectively, and the spectrum of a random coil chain: less negative ellipticity at 222 nm, less positive ellipticity at 190 nm, and a shift to lower wavelengths of the D1 minimum at 208 nm and the D3 minimum at 202 nm. These results suggest that D2 behaves as a disordered random coil, as predicted. Furthermore, the spectrum of D1D2D3 indicates that the secondary structure of ING4 is essentially the same as the sum of the secondary structures of its three regions, suggesting that each domain is structurally independent from the rest of the chain.

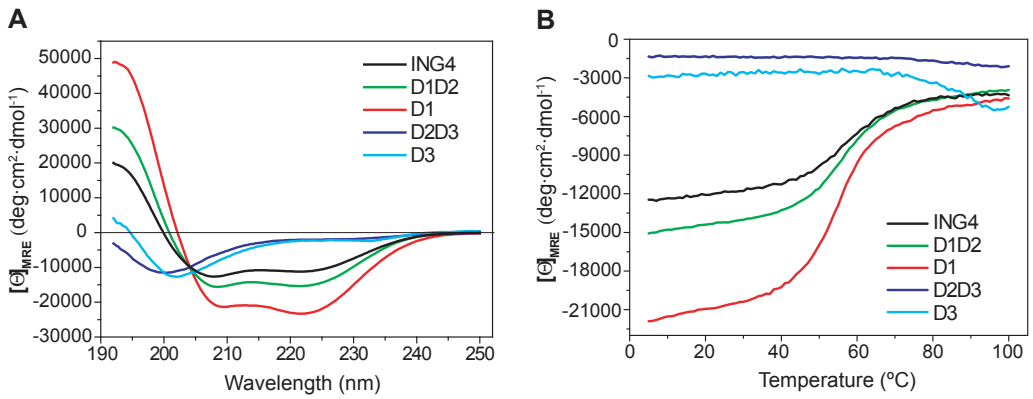

Figure 4. Structure and stability of ING4 analysed by circular dichroism. A. Far-UV spectra of the five constructs recorded at 25 ºC in 20mM sodium phosphate buffer pH 6.5, 200 mM NaCl, 1mM DTT. B. Thermal denaturation curves of the same molecules and in the same buffer as in A followed by the change in their molar residue ellipticity at 222 nm.

The stability and unfolding cooperativity of the different constructs was evaluated by thermal denaturation CD measurement at 222 nm. The transition curves show cooperative unfolding transitions for the full-length protein, D1 and D1D2 constructs (with similar melting temperatures around 55 ºC, as well as for D3 (mid point denaturation temperature around 90 ºC). The strongest signal in the protein comes from the helical D1 domain and thus it dominates the transition in those constructs containing it, while the random coil D2 region contributes a nearly invariant molar ellipticity that flatens the curves of the corresponding constructs and dominates that of D2D3 construct (still when expanded in the vertical scale a transition similar to that of D3 can be observed). These results indicate that D1 and D3 are domains with well defined tertiary structures while D2 behaves like a random coil and that the three domains are independent with weak interactions among them, if any.

The Kratky representation of the scattering profiles has been used to discriminate between folded and unfolded proteins (47). The Kratky plot of D1 (supplementary figure 3) display a pronounced peak that indicates that it is a folded protein. Conversely, D1D2 Kratky plot presents a dual behaviour where the presence of a less pronounced peak is added to the continuous increase of the $I(s)s^2$ values with s that is typical of unstructured systems. D1D2 also presents large values of R_g and D_{max} (Table 1) that are in agreement with the qualitative Kratly representation. These results obtained by SAXS data are consistent with the sequence analysis, CD, and thermal denaturation.

The two-dimensional ^{1}H-^{15}N-HSQC NMR spectra of ING4 and the different constructs registered identical conditions were compared regarding the number of signals, as well as their width and dispersion (Figure 5 and Supplemental figure 1). The spectrum of ING4 shows a mixture of dispersed and non-dispersed signals, and much narrower than would correspond to a dimeric molecule of 57 kD. This result is in agreement with the existence of well folded D1 and D3 linked by a flexible random coil D2, which will made the tumbling of the folded domains largely independent of the rest of the molecule. The spectrum of D1 (1-118) spectrum shows less peaks than the expected from its length and were mostly narrow and with low dispersion. This result is consistent with a 28 kDa dimer folded into an elongated coiled coil helical structure. This molecular size, shape and the alignment of the N-H bond with the α-helix axis result altogether in very unfavourable nuclei relaxation properties and only the most flexible parts of the molecule are observed in these conditions. At higher temperatures, but below the mid denaturation point, a larger number of dispersed peaks are visible (suppl. Fig. 2) due to increased global mobility. D3, whose spectrum has already been assigned (12) corresponds to a small globularly folded protein. In the spectra of D1D2 and D2D3 a larger number of non-dispersed peaks are observed as expected from a random coil chain extension. The superimposition of the spectra of D1D2 with D3 or D1 with D2D3 (Fig. 5) results in a nearly identical pattern of signals and very similar to that of full length ING4 of the NMR spectra as corresponds to essentially independent and non-interacting domains[1].

3.5 The N-terminal domain of ING4 is an elongated coiled coil structure.

While the circular dichroism spectrum shows that D1 contains helical structure organized into coiled coils, a glimpse of its overall shape can be obtained from small angle X-ray scattering measurements. The distance distribution function, *p(r)*, of D1 (supplementary Figure 3) was calculated using a

[1] A larger number of narrow signals are observed in D2D3 than in D1D2, which may be result from interactions of the D1 domain with nearby residues of the D2 domain making them invisible. Some ordering in this region close to the D1 could be responsible for the slightly larger melting temperature of D1D2 than D1 observed by CD.

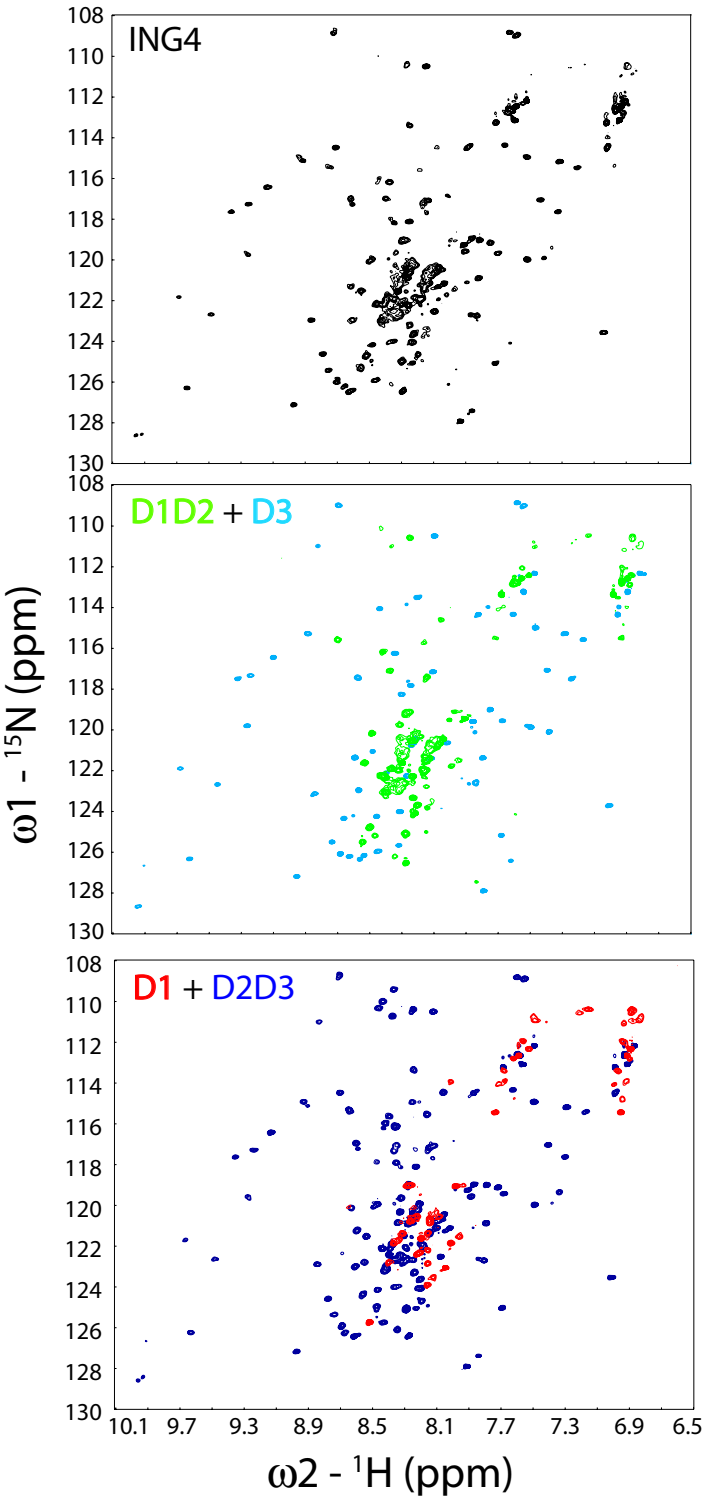

Figure 5. ^{1}H-^{15}N HSQC NMR spectra of ING4 and their domains. Conditions: 20 mM sodium phosphate, 50 mM NaCl 25 ºC. The spectra have been processed in the same way and plotted to show signals just above the overall noise.

maximum dimension of 104 Å using the program GNOM. It presents a peak position around 20 Å, and the function decreases linearly towards D_{max}, although the slope smoothes at the largest distances. In a first approximation, these features are characteristic of a *p(r)* belonging to an elongated cylinder with an approximate diameter of 40 Å (48, 49). With an axial rise of 1.5 Å/residue the length is slightly larger than expected for a 50 residue long helix (75 Å) but smaller that for a 100 residue long one (150 Å). The radius of the cylinder, around 40 Å, is slightly larger than expected for an assembly of four helices (50). We interpret these data as the D1 domain of ING4 folding into a helix loop helix structure which dimerizes into a four helix-coiled coil. The longer measured length could be due either to the chain ends of D1 (that may possess some degree of disorder, as suggested by the NMR spectrum) or to a dimerization that is not in register but shifted so that the overall length of the coiled coil is larger than the length of the individually packed helices. A slightly curved assembly could be the origin of the broad diameter of the cylindrical approximation (51). A more precise picture of the overall shape of D1 was obtained by the *ab initio* reconstruction of D1 which displays a cylindrical elongated shape with a certain degree of bending (supplementary figure 4). Unfortunately, we can not derive from these data how the dimers and helices interact with each other (it could be parallel or antiparallel). Helical wheel analysis of the different arrangements failed to show a clear preference for any of them based on the pattern of buried and exposed amino acids (data not shown). Unfortunately, the high mobility of D1D2 precluded a similar structural modelling for this construction.

3.5 A direct interaction between ING4 and p53 is not confirmed by NMR.

The tumour suppressor activity of ING4 has been described as being p53 dependent (2), and a physical interaction has been observed by immunoprecipitation (17) and also by pull-downs with the isolated p53 and GST fusions of different ING4 constructs (26). Positive results were obtained with those constructs containing the NLS region, one of them spanning residues 120-248, almost identical to our D2D3 construct. We have tried to investigate if this interaction could be observed and characterised in detail by

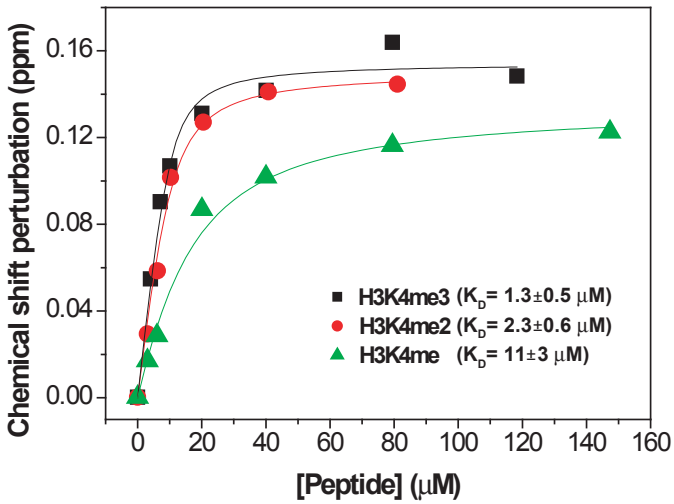

Figure 6. ING4 binding to histone 3 methylated peptides by NMR. The chemical shift perturbation experienced by the signal of W237 signal in the HSQC spectrum of full-length ING4 is represented as a function of added peptide.

NMR. For that purpose, ^{1}H-^{15}N-HSQC spectra of D2D3 were recorded in the presence and the absence of p53 (suppl. Fig 5) at a 1:1 monomer:monomer ratio. No evidence of interaction was observed. Spectra of ^{15}N-^{2}H-labelled p53 in the presence of unlabelled D2D2 was also indicative of the absence of the interaction, and the same result was obtained when full length ING4 was added to labelled p53. These results show there is no direct interaction between ING4 and p53 contrary to what have been previously reported (26) and suggest that other molecules mediate the interaction of ING4 with p53.

3.6 ING4 recognition of histone 3 methylated at Lys4.

We have reported that the PHD finger of ING4 (D3) specifically binds to H3K4mex (X=3,2,1) with a dissociation constant (K_D) in the μM range (12, 13). The observation of a set of NMR signals in the ING4 spectrum that closely matches those assigned in the spectrum of the isolated PHD allows us to study this recognition process in the context of the full ING4 dimeric molecule. Upon addition of the trimethylated histone peptide (H3K4me3), the pattern of signals corresponding the D3 domain experience changes that are qualitatively and quantitatively nearly identical to those observed when the isolated PHD domain is studied (data not shown). A titration of ING4 using as probe the signal of W237 (assigned base on the similarity of its resonance frequencies with those of the isolated PHD finger) yields a $K_D = 1.3 \pm 0.5$ μM) very similar to that of the PHD alone ($K_D = 3.9 \pm 0.7$ μM). Similar affinities are measured for the mono- and di-methylated forms of the histone, demonstrating that ING4 hardly discriminates among the three metylated states, again as previously found for the isolated PHD (Figure 6).

4. Conclusions.

The tumor suppressor ING4 is structurally organized into three independent domains: a helical coiled coil N-terminal domain, a central flexible region and a PHD finger. ING4 forms dimers *in vitro* and inside living cells through its N-terminal domain. It does not interact with p53 through the central region, and binds methylated histone 3 methylated at K4 in the same mode and with the same affinity as its isolated PHD. Therefore ING4 is a bivalent reader of this post-translational histone modification. The N-terminal region of ING4 contains a conserved PCR (Potential Chromatin Regulatory) motif which could also be the site of interaction with components of the histone acetyl transferase complex HBO1. In particular, the protein hEaf6 contains a sequence with a high potential to form a coiled-coil structure which could assemble with the PCR of ING4, recruiting the HBO1 complex to chromatin sites with the H3K4me3 modification.

Acknowledgements

We thank Daniel Padró and Stefan M. Freund, for help with the NMR experiments, Lei Cheng, Cintia González and Carolie M. Blair for help with protein purification. This work was supported by grant CTQ2008-03115/BQU from the Ministerio de Ciencia en Innovación, EU-grant 3D-REPERTOIRE, contract no. LSHG-CT-2005-512028, and a grant from Fundación Médica Mutua Madrileña. P.B. holds a Ramón y Cajal contract. We acknowledge the support of the European Community – Research infrastructure Action under the FP6 "Structuring the European research area program contract number RII/2004/5060008" to the EMBL-Hamburg outstation that covered the travel and accommodation expenses at the EMBL-Hamburg.

References.

1. Campos, E. I., Chin, M. Y., Kuo, W. H., and Li, G. Biological functions of the ING family tumor suppressors. Cell Mol Life Sci, 61: 2597-2613., 2004.
2. Gong, W., Suzuki, K., Russell, M., and Riabowol, K. Function of the ING family of PHD proteins in cancer. Int J Biochem Cell Biol, 37: 1054-1065, 2005.
3. Shi, X. and Gozani, O. The fellowships of the INGs. J Cell Biochem, 2005.
4. Soliman, M. A. and Riabowol, K. After a decade of study-ING, a PHD for a versatile family of proteins. Trends Biochem Sci, 32: 509-519, 2007.
5. Garkavtsev, I., Kazarov, A., Gudkov, A., and Riabowol, K. Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. Nat Genet, 14: 415-420., 1996.
6. He, G. H., Helbing, C. C., Wagner, M. J., Sensen, C. W., and Riabowol, K. Phylogenetic analysis of the ING family of PHD finger proteins. Mol Biol Evol, 22: 104-116, 2005.
7. Doyon, Y., Cayrou, C., Ullah, M., Landry, A. J., Cote, V., Selleck, W., Lane, W. S., Tan, S., Yang, X. J., and Cote, J. ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell, 21: 51-64, 2006.
8. Gozani, O., Karuman, P., Jones, D. R., Ivanov, D., Cha, J., Lugovskoy, A. A., Baird, C. L., Zhu, H., Field, S. J., Lessnick, S. L., Villasenor, J., Mehrotra, B., Chen, J., Rao, V. R., Brugge, J. S., Ferguson, C. G., Payrastre, B., Myszka, D. G., Cantley, L. C., Wagner, G., Divecha, N., Prestwich, G. D., and Yuan, J. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell, 114: 99-111., 2003.
9. Pena, P. V., Hom, R. A., Hung, T., Lin, H., Kuo, A. J., Wong, R. P., Subach, O. M., Champagne, K. S., Zhao, R., Verkhusha, V. V., Li, G., Gozani, O., and Kutateladze, T. G. Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. J Mol Biol, 380: 303-312, 2008.
10. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature, 442: 100-103, 2006.
11. Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G., Shi, Y., Cote, J., Chua, K. F., and Gozani, O. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature, 442: 96-99, 2006.
12. Palacios, A., Garcia, P., Padro, D., Lopez-Hernandez, E., Martin, I., and Blanco, F. J. Solution structure and NMR characterization of the binding to methylated histone tails of the plant homeodomain finger of the tumour suppressor ING4. FEBS Lett, 580: 6903-6908, 2006.
13. Palacios, A., Munoz, I. G., Pantoja-Uceda, D., Marcaida, M. J., Torres, D., Martin-Garcia, J. M., Luque, I., Montoya, G., and Blanco, F. J. Molecular basis of histone H3K4me3 recognition by ING4. J Biol Chem, 283: 15956-15964, 2008.
14. Champagne, K. S., Saksouk, N., Pena, P. V., Johnson, K., Ullah, M., Yang, X. J., Cote, J., and Kutateladze, T. G. The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. Proteins, 72: 1371-1376, 2008.

15. Pascual, J., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. J Mol Biol, 304: 723-729., 2000.
16. Aasland, R., Gibson, T. J., and Stewart, A. F. The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci, 20: 56-59., 1995.
17. Shiseki, M., Nagashima, M., Pedeux, R. M., Kitahama-Shiseki, M., Miura, K., Okamura, S., Onogi, H., Higashimoto, Y., Appella, E., Yokota, J., and Harris, C. C. p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity. Cancer Res, 63: 2373-2378., 2003.
18. Garkavtsev, I., Kozin, S. V., Chernova, O., Xu, L., Winkler, F., Brown, E., Barnett, G. H., and Jain, R. K. The candidate tumour suppressor protein ING4 regulates brain tumour growth and angiogenesis. Nature, 428: 328-332., 2004.
19. Unoki, M., Shen, J. C., Zheng, Z. M., and Harris, C. C. Novel splice variants of ING4 and their possible roles in the regulation of cell growth and motility. J Biol Chem, 281: 34677-34686, 2006.
20. Shen, J. C., Unoki, M., Ythier, D., Duperray, A., Varticovski, L., Kumamoto, K., Pedeux, R., and Harris, C. C. Inhibitor of growth 4 suppresses cell spreading and cell migration by interacting with a novel binding partner, liprin alpha1. Cancer Res, 67: 2552-2558, 2007.
21. Colla, S., Tagliaferri, S., Morandi, F., Lunghi, P., Donofrio, G., Martorana, D., Mancini, C., Lazzaretti, M., Mazzera, L., Ravanetti, L., Bonomini, S., Ferrari, L., Miranda, C., Ladetto, M., Neri, T. M., Neri, A., Greco, A., Mangoni, M., Bonati, A., Rizzoli, V., and Giuliani, N. The new tumor-suppressor gene inhibitor of growth family member 4 (ING4) regulates the production of proangiogenic molecules by myeloma cells and suppresses hypoxia-inducible factor-1 alpha (HIF-1alpha) activity: involvement in myeloma-induced angiogenesis. Blood, 110: 4464-4475, 2007.
22. Susan Nozell, T. L., Dorothy Moseley, Lisa Nowoslawski, Marijke DeVos, George P. Atkinson, Keith Harrison, L. Burton Nabors, and Etty N. Benveniste The ING4 Tumor Suppressor Attenuates NF-B Activity at the Promoter of Target Genes 2008 (published online).
23. Kim, S., Chin, K., Gray, J. W., and Bishop, J. M. A screen for genes that suppress loss of contact inhibition: identification of ING4 as a candidate tumor suppressor gene in human cancer. Proc Natl Acad Sci U S A, 101: 16251-16256. Epub 12004 Nov 16204., 2004.
24. Kim, S. HuntING4 new tumor suppressors. Cell Cycle, 4: 516-517, 2005.
25. Zhang, X., Xu, L. S., Wang, Z. Q., Wang, K. S., Li, N., Cheng, Z. H., Huang, S. Z., Wei, D. Z., and Han, Z. G. ING4 induces G2/M cell cycle arrest and enhances the chemosensitivity to DNA-damage agents in HepG2 cells. FEBS Lett, 570: 7-12., 2004.
26. Zhang, X., Wang, K. S., Wang, Z. Q., Xu, L. S., Wang, Q. W., Chen, F., Wei, D. Z., and Han, Z. G. Nuclear localization signal of ING4 plays a key role in its binding to p53. Biochem Biophys Res Commun, 331: 1032-1038, 2005.
27. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, 25: 4876-4882, 1997.
28. Feng, X., Hara, Y., and Riabowol, K. Different HATS of the ING1 gene family. Trends Cell Biol, 12: 532-538, 2002.

29. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., and Barton, G. J. JPred: a consensus secondary structure prediction server. Bioinformatics, 14: 892-893, 1998.
30. Cornilescu, G., Delaglio, F., and Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR, 13: 289-302, 1999.
31. Veprintsev, D. B., Freund, S. M., Andreeva, A., Rutledge, S. E., Tidow, H., Canadillas, J. M., Blair, C. M., and Fersht, A. R. Core domain interactions in full-length p53 in solution. Proc Natl Acad Sci U S A, 103: 2115-2119, 2006.
32. Laue, T. M. S., Shah,B.D., Ridgeway,T.M. and Pelletier,S.L. Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins. Royal Society of Chemistry, Cambridge, UK., 1992.
33. Schuck, P. and Rossmanith, P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. Biopolymers, 54: 328-341, 2000.
34. Minton, A. P. Modern Analytical Ultracentrifugation. Birkhauser Boston, Inc., Cambridge, MA., 1994.
35. Chen, Y. H., Yang, J. T., and Chau, K. H. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. Biochemistry, 13: 3350-3359, 1974.
36. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR, 6: 135-140, 1995.
37. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR, 6: 277-293, 1995.
38. González, L., Freije, J. M., Cal, S., López-Otín, C., Serrano, M., and Palmero, I. A functional link between the tumour suppressors ARF and p33ING1. Oncogene: in press, 2006.
39. Palmero, I., Murga, M., Zubiaga, A., and Serrano, M. Activation of ARF by oncogenic stress in mouse fibroblasts is independent of E2F1 and E2F2. Oncogene, 21: 2939-2947, 2002.
40. Roessle, M. W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C. & Svergun, D.I. Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. Journal of applied cristallography, 40: 190-194, 2007.
41. Petr V. Konarev, V. V. V., Anna V. Sokolova, Michel H. J. Koch and and Svergun, D. I. PRIMUS: aWindows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography, 36: 1277–1282, 2003.
42. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of applied crystallography, 25: 495-503, 1992.
43. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. Biophys J, 80: 2946-2953, 2001.
44. Volkov, V. V. S., D. I. Uniqueness of ab initio shape determination in small-angle scattering. . Appl. Crystallogr, 36: 860-864, 2003.
45. Lupas, A., Van Dyke, M., and Stock, J. Predicting coiled coils from protein sequences. Science, 252: 1162-1164, 1991.

46. Dutta, K., Alexandrov, A., Huang, H., and Pascal, S. M. pH-induced folding of an apoptotic coiled coil. Protein Sci, 10: 2531-2540, 2001.
47. Doniach, S. Changes in biomolecular conformation seen by small angle X-ray scattering. Chem Rev, 101: 1763-1778, 2001.
48. Glatter, O. K. a. O. Small angle x-ray scattering. 1982.
49. Svergun, M. H. J. K. a. D. I. Small-angle scattering studies of biological macromolecules in solution. Rep. Prog. Phys. , 66 1735-1782 2003
50. Tarbouriech, N., Curran, J., Ebel, C., Ruigrok, R. W., and Burmeister, W. P. On the domain structure and the polymerization state of the sendai virus P protein. Virology, 266: 99-109, 2000.
51. Yousef, M. S., Kamikubo, H., Kataoka, M., Kato, R., and Wakatsuki, S. Miranda cargo-binding domain forms an elongated coiled-coil homodimer in solution: implications for asymmetric cell division in Drosophila. Protein Sci, 17: 908-917, 2008.

Suplementary material.

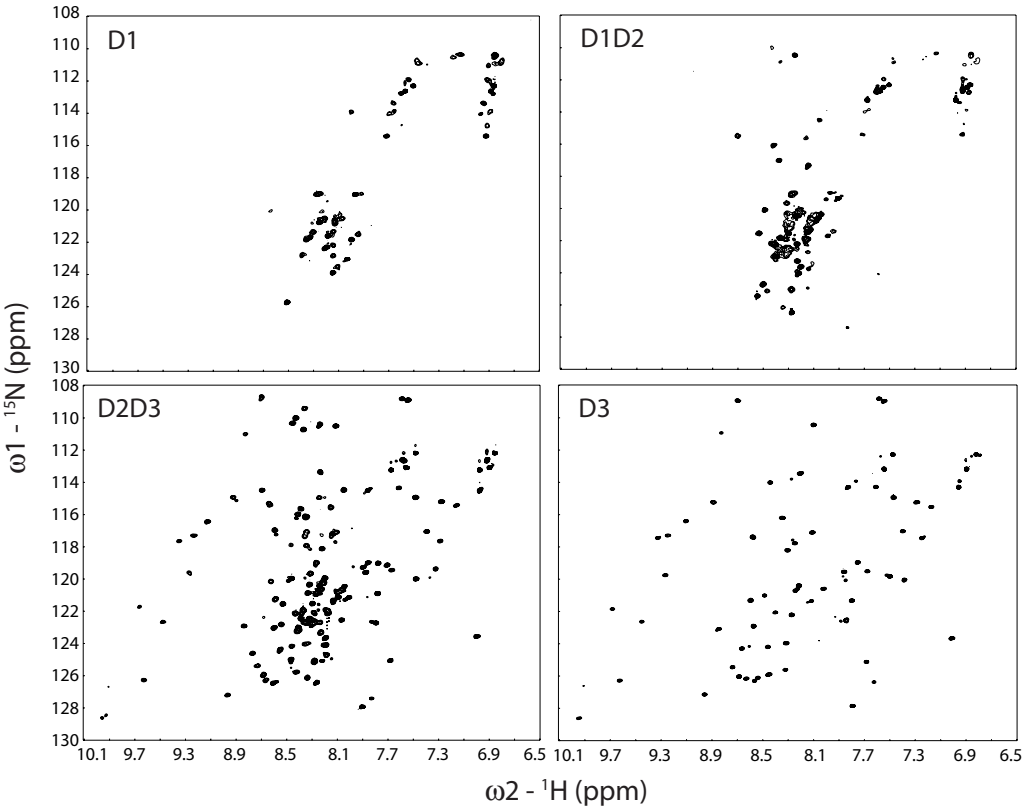

Figure 1. ^{1}H-^{15}N HSQC spectra of the different constructs in 20 mM sodium phosphate, 50 mM NaCl 25 ºC. The spectra have been processed in the same way and plotted to show signals just above the overall noise.

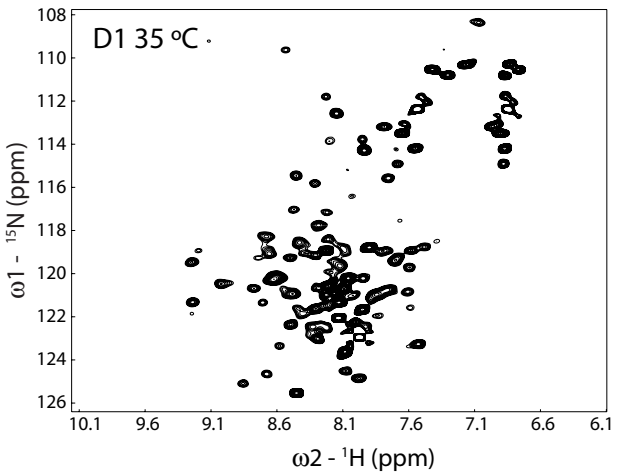

Figure 2. ^{1}H-^{15}N HSQC spectrum of D1 at 35 ºC in 20 mM sodium phosphate pH 6.5, 200 mM NaCl 1mM perdeuterated dithiothreitol (d_{10}-DTT), 5% $^{2}H_2O$ and 0.01% NaN_3.

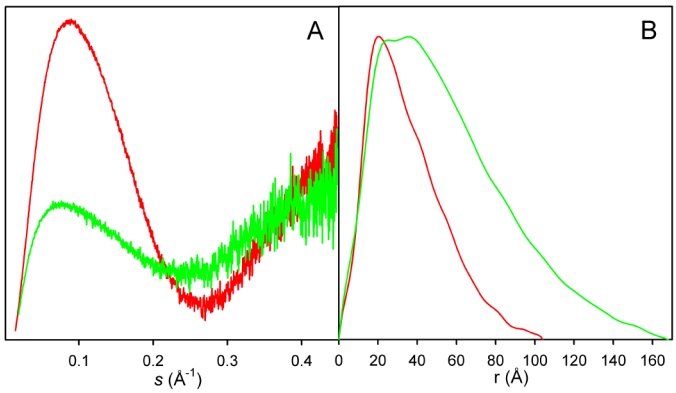

Figure 3. A. Kratky plots of D1 (red) and D1D2 (green). B. Distance distribution function, p(r), of D1 (red) and D1D2 (green).

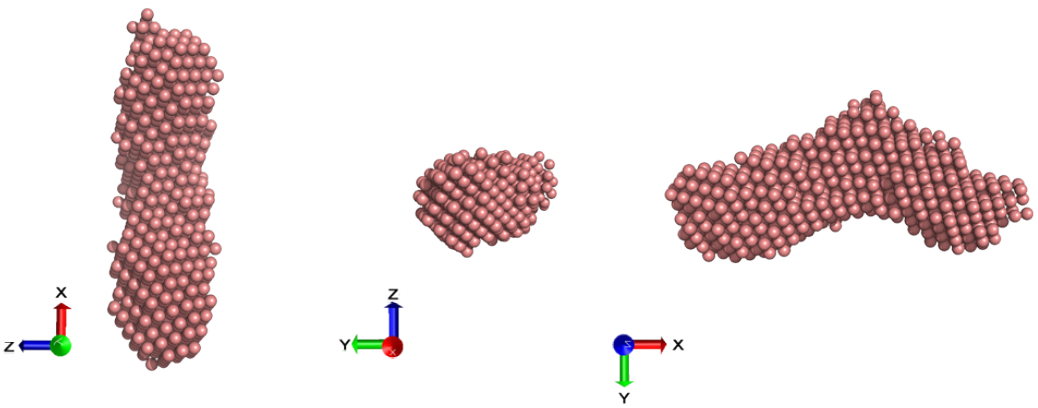

Figure 4. Low resolution ab initio SAXS model of D1 computed using the program GASBOR. Three different views along each of the orthogonal axis are shown.

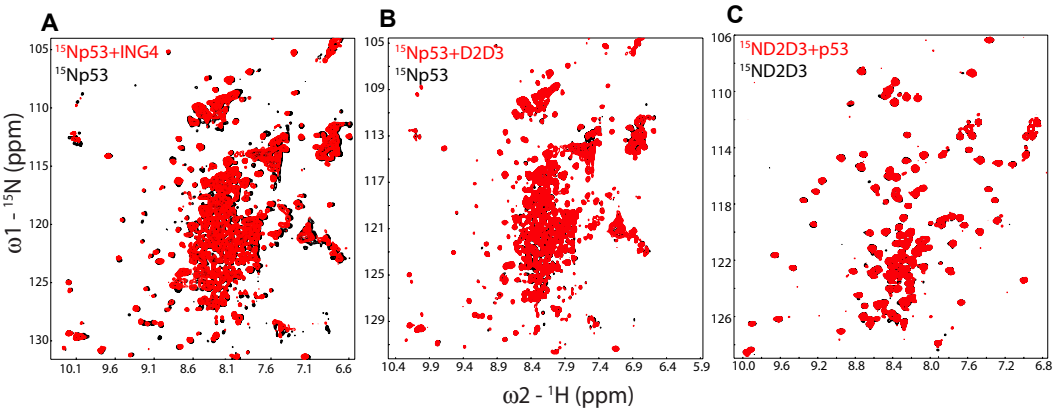

Figure 5. Overlays of ^{1}H-^{15}N-HSQC spectra of 1:1 (monomer molar ratio) mixtures of p53 and ING4 or D2D3 in 25 mM sodium phosphate pH 7.2, 150 mM NaCl, 5 mM perdeuterated DTT. Left: perdeuterated U-^{15}N p53 plus unlabelled ING4 (red) overlaid on perdeuterated U-^{15}N-p53 (black). Middle: perdeuterated U-^{15}N-p53 plus unlabelled D2D3 (red) overlaid on perdeuterated U-^{15}N p53 (black). Right: U-^{15}N-ING4 D2D3 plus unlabelled p53 (red) overlaid on U-^{15}N-D2D3 (black).

Appendix: Visualization of ING4 homodimerization in living cells by using Bimolecular Fluorescence Complementation (BiFC) Analysis

Abstract

The ING4 nuclear protein belongs to the ING (Inhibitor of growth) tumor suppressor family. Sequence homology and phylogenetics indicate that ING proteins have a similar architecture, with conserved N- and C-terminal regions, and a central non conserved region with the nuclear localization signal. ING4 is a dimer in vitro, as a pure protein in solution and also inside transfected cells, as seen by inmunoprecipitation experiments. The dimerization site is at the N-terminal domain, which is rich in helical structure, likely with a coiled coil structure could also be the interact site with some components of the histone acetylation complex HBO1, participating in regulation of gene transcription and. The C-terminal plant homeodomain (PHD) constitutes the site for recognition of histone 3 methylated at residue Lys4. ING4 would bind trimethylated histone 3 tails working as a bivalent reader of the H3K4me3 histone modification recruiting HBO1 (and perhaps other chromatin remodeling complexes to chromatin location enriched in this histone modification. We set out to directly visualize interactions among monomers of ING4 proteins in order to elucidate the selectivity of dimmer formation in living cells by bimolecular fluorescence complementation experiments. The results show how BiFC result as an unspecific assay in the context of ING4, probably must to ING4 is part of cellular complexes inside the nucleus.

Key words: ING4, inhibitor of growth, PHD, tumour suppressor, BiFC, histone, chromatin remodeling complexes.

Abbreviations: ING, inhibitor of growth; NLS, nuclear localization signal; PHD, plant homeodomain; LZL, Leucine zipper-like motif; BiFC, Bimolecular fluorescence complementation; VFP, venus fluorescent protein; CFP, cyan fluorescent protein; YFP, yellow fluorescent protein; MCS, multiple cloning site; TSG, tumor suppressor gene; PCNA, proliferating cell nuclear antigen; PBR, polybasic region; NTS, nuclear translocation signals; NCR, novel conserved region; HAT, histone acetylated complexes; HDAC, histone desacetylated complexes; NF-kB, nuclear factor kappa-light-chain-enhancer of activated B cells; HIF-1, hypoxia inducible factor-1; H3K4me3, trimethylated histone 3 at lysine 4; VN, Venus (1-173); VC, Venus (155-238).

1. Introduction

The ING proteins are involved in multiple cellular processes implicated in growth regulation, senescence, DNA repair, oncogenesis, apoptosis, cell migration and angiogenesis (1-4). The inhibitor of growth family of proteins is an evolutionary conserved family frequently deregulated in different cancer types and present from yeast to humans (1,2,5). ING1, the founding member, was described in 1996 (6) and classified as tumor suppressor gene (TSG) type II. Subsequently, 4 new members of the ING family were identified by homology search: ING2-ING5. Phylogeny studies have found that ING1 and ING2 proteins on one hand and ING3-ING5 proteins in the other hand have an overall high homology and therefore, could have related functions (7,8). Bioinformatic analyses of ING proteins revealed several conserved regions, most of which have been linked to specific functions (5). Some of these regions have

been found in a subset of the ING proteins, such as the PCNA-interacting-protein motif which mediates the interaction between ING1b and PCNA in a DNA damage inducible manner (9). Human ING2-ING5 contains in their N-terminal domain a LZL motif, which in ING2 seems affect its function in DNA repair and apoptosis (10). The central region of the ING proteins has a nuclear localization sequence (NLS) that contain distinct nucleolar translocation sequences (NTS) (5). At the end of C-terminus domain of ING1 and ING2 is found a short region called the polybasic region (PBR) that bind phosphatidylinositol monophosphates, being these ING proteins defined as nuclear receptor of phosphoinositides (11-13).

In contrast to these motifs, all ING from all species share a PHD finger that contains a C4HC3 zinc finger motif (four cysteines, one histidine and three cysteines with characteristic spacing) located near the C-terminus (14). The PHD finger has recently been shown to interact specifically with methylated forms of histone3, particularly those trimethylated at lysine 4 (4,15-18). The next best conserved structural feature, called the potential chromatin regulatory (PCR), is found at N-terminal domain and might participate in binding to HAT or HDAC chromatin remodelling complexes (5,19). Thus, ING proteins are thought to function as chromatin regulatory molecules, acting as co-factors for distinct histones and acetyl-transferase (HAT) and deacetylase (HDAC) enzyme complexes.

After ING1, ING4 is the most investigated ING gene in tumors. The ING4 gene, located at chromosome 12p13.31, consists of 8 exons and encodes a 29-kDa protein (5,20,21). ING4 is a nuclear factor expressed in all normal tissues and markedly reduced in glioblastoma cells, and head and neck sqamous cell carcinoma with levels correlates with the progression from lower to higher grades of tumour (20,21). ING4 is involved in regulating brain tumour growth and angiogenesis through their association with NF-κB, being this activity of ING4 unique in the family. ING4 physically interacts with p65 (RelA) subunit of nuclear factor NF-κB, forming a transcripcional complex, and represses NF-κB responsive genes (20). Moreover ING4 regulate the activity of hypoxia inducible factor (HIF)-1 α transcription factor under hypoxic conditions (22).

ING4 was also detected in a search for genes that suppressed the loss of contact inhibition, a characteristic in vitro property of cancerous cells (23). This and other studies suggested that ING4 suppresses tumour biology without directly deregulating the cell cycle (24). Different mutations in the ING4 transcript were detected in several cancer lines, and a loss of function has been proposed for all of them (23).

ING4 regulate gene expression by localizing histone-modifying enzymes to sites of active transcription and interacting with various transcription factors, since ING4 protein was found associated to the HAT complex containing HBO1, hEafS and JADE1/2/3 proteins (25-27). Moreover, we have reported the molecular basis of the H3K4me3 recognition by the PHD finger of ING4 (16) where ING4 discriminate between methylated versus non-methylated histone tails but there is little if any differentiation among the three possible methylated states. Then, ING4 could function in a combinatorial way, first recognizing and binding the methylated histones trough its PHD domain, and after recruiting chromatin remodelling complexes (HAT) to them trough it N-terminal domain, with the aim of regulate gene expression.

Recently, we have described experimentally the overall structural organization of ING4, showing that is a dimer in vitro, as a pure protein in solution (analytical ultracentrifugation) and also inside transfected cells (by pulldown assays). The dimerization site is at the N-terminal domain, which is rich in helical structure, likely with a coiled coil structure, and contains the LZL and PCR motives. The central domain

is a non conserved and mostly unfolded region and the C-terminal domain constitutes de Plant Homeo Domain finger.

We propose that dimeric ING4 could function as a bivalent reader of histone code markers and, also, in the recognition of macromolecular complexes involves in transcription regulation. To confirm this interaction in vivo, we used bimolecular fluorescence complementary analysis. BiFC allow directly visualize interactions among monomers of ING4 in living cells with minimal perturbation of the normal cellular environment.

2. Materials and Methods

2.1 Construction of plasmid vectors.

BiFC constructs pBiFC-bJun-VN173C (pFlag-CMV backbone), pBiFC-bFos-VC155C (pCMV-HA backbone), pBiFC-bFos(ΔZip)-VC155C (pCMV-HA backbone), pBiFC-VN173N (pFlag-CMV backbone), pBiFc-VC155N (pCMV-HA backbone) were transferred by Chang-Deng Hu M.D. laboratory (28). Plasmids for the analysis of ING4 dimerization were constructed by cloning the coding regions of ING4full-length (1-249) into pBiFC-VN173N (MCS at C termini of Venus) or pBiFC-VN173C (MCS at N termini of Venus) and into pBiFc-VC155N (MCS at C termini of Venus) or pBiFC-VC155C (MCS at N termini of Venus). In this way, ING4 was cloned after or before coding region of Venus fluorescence fragment (29) with a linker between them of GGGGSGGGGS, RSIAT or RPACKIPNDLKQKVMNH amino acid sequence in the cases where ING4 there was been cloned after Venus sequence (into pBiFC-VN173N and pBiFC-VC155N), into pBiFC-VN173C or into pBiFC-VC155C, respectively. ING4Δ(1-118) (residues 119-249 of ING4) were subcloned in the same manner that ING4 full-length.

2.2 Western blot assay.

To determine expression levels of the fusion proteins, 293T cells were grown in 6 well culture plates with DMEM supplemented with FBS and antibiotics during 24 h, transfected with pBiFC-bJun-VN173 (pFlag-CMV backbone), pBiFC-bFos-VC155 (pCMV-HA backbone), pBiFC-bFosΔZip-VC155 (pCMV-HA backbone), pBiFC-VN173N-ING4 (pFlag-CMV backbone), pBiFc-VC155N-ING4 (pCMV-HA backbone), pBiFC-VN173N-ING4Δd1 (pFlag-CMV backbone) and pBiFc-VC155N-ING4Δd1 (pCMV-HA backbone) plasmids, lysed with cell scrapped in lysis buffer containing protease inhibitors and harvested 48 hours after transfection. The supernatants of lysates were charged on a PAGE-SDS and analyzed by western blot analysis using anti-HA antibody (Roche) or anti-Flag antibody (Sigma). Proteins were detected using HRP-conjugated antimouse secondary antibody (Dako) and ECL reagents (Amersham Biosciences).

2.3 Inmunofluorescence.

HEK293T cells used to BiFC analysis were grown for 24 h on chamber slides (Lab-Tek™ chamber slide, Nunc) previously treated with IV collagen. The slides were washed with phosphate buffer saline (PBS), fixed with PBS containing 4 % formaldehyde, permeabilized with PBS containing 0.1 % triton X-100, blocked with 2% BSA in PBS at 4ºC overnight, incubated witha primary mouse antibody against HA (Roche) or Flag (Sigma) 1h at room temperature, incubated with secondary ALEXA-fluor conjugated

goat anti-mousse antibody (Invitrogen) during 30min at room temperature and finally mounted with Vectashield containing DAPI (1:2). The images were acquired with an inverted fluorescence microscope (Axiophot Zeiss) with filters designed for the detection of green light.

2.4 Fluorescence imaging of living cells.

For BiFC analysis, 293T Cells were maintained in DMEM supplemented with 10% fetal bovine serum (Gibco) and 5 mg/mL streptomycin (Cepa) and 5000 units/mL penicillin (Level). For transient transfections, 2.75 x 10^4 cells were platted into 8 well glass bottom dish (LabTek II, Nunc) 24 h before transfections. To control and optimize the expression level, the cells were transfected with different amounts of plasmid DNA and observed at different times after transfection. Finally, cells were cotransfected with plasmids (0.1 μg of each DNA) encoding ING4/CFP proteins, ING4 Δ(1-118)/CFP proteins, bFos/bJun/CFP proteins and bFosΔZip/bJun/CFP proteins indicated with Fugene 6 (Roche Diagnostics) as the transfection agent. After 24 h, images were acquired with filters optimized for YFP and CFP fluorescence in a LSM 510 META confocal microscope (Zeiss).

2.5 Quantification of fluorescence complementation efficiency.

The fluorescence intensities produced by bimolecular fluorescence complementation (VN-VC) and the internal reference (CFP) were measured integrating the area of each cell (at least 50) using filters optimized for the separation of CFP and YFP (Venus). The ratios of YFP/CFP fluorescence for each individual cell were calculated getting a distribution of the ratios.

3. Results and discussion

3.1 Direct visualization of ING4 dimerization.

To determine if ING4 could form homodimers inside the cells, we used bimolecular fluorescence complementation (BiFC) analysis to visualize dimerization in living cells. The BiFC approach is based on the formation of a bimolecular fluorescent complex when two non-fluorescent fragments of Venus

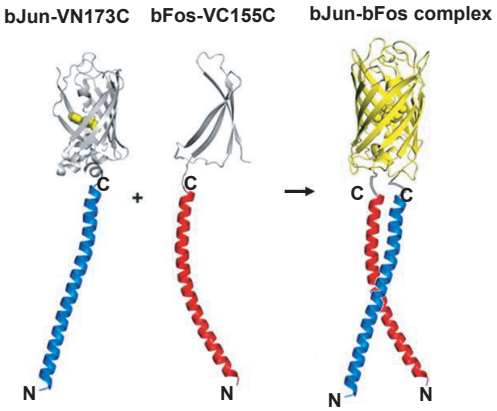

Figure 1. Formation of the bFosVC155C-bJunVN173C bimolecular fluorescence complex. Hypothetical model of the interaction between bJun (blue) and bFos (red) brings N- and C-terminal fragments of Venus into proximity to form a BiFC complex.

(VN and VC) (Fig 1), are brought together by an interaction between proteins fused to the fragments (30). In order to design BiFC fusion constructs we had to consider some critical parameters as fragments of fluorescent proteins, positions of fusions and the linkers. Fragments of Venus, a variant of yellow fluorescent protein (29), truncated at residue 155 and 173 (designated VN173 containing N-terminal residues 1 to 173, and VC155, containing residues 155 to 238) have been recommended for most purposes (28), due to their high signal output and specificity. The position of fusions was determined considering our *in vitro* structural studies about ING4 where we saw that the N-terminal domain constitutes the dimerization domain. Therefore, we designed fusion constructs with Venus at the N-termini of ING4, but also the opposite combination, so that we could test the different combinations. To facilitate the detection of their expression, the fragments fused to VN173 have a FLAG® tag at their N termini, and the fragments fused to VC155 have a hemaglutinin (HA) tag at their N termini. To connect the fragments of the fluorescent proteins to ING4 in the fusion proteins, we used the linkers recommended for the author of this method (31).

In order to confirm that the fusions do not affect the localization or the stabilities of the ING4, we determined by indirect inmunofluorescence (data not shown) that the fusions maintain the nuclear localization of ING4 and ING4Δ(1–118), by inmunoblot analyses we confirm a high level of expression similar for each fusion. So, we have managed to keep the localization and the stability of the ING4 in the fusions.

It is known that the fluorescent protein fragments are able to form fluorescent complexes with a low efficiency even in the absence of a specific interaction. So, in order to determine if any fluorescence observed reflects a specific protein interaction, it is essential to include positive and negative controls in each BiFC experiment.

First of all we use a model system of interaction studied with BiFC (28,32) as positive (VN173C-bJun and VC155C-bFos) and negative (VN173C-bJun and VC155C-bFos(ΔZIP)) controls of the experiment. Both combinations exhibit fluorescence (Fig 2, A and B), but the pattern is very different, being uniform fluorescence at nucleus in the case of the positive control and fluorescence only centered in nucleolus in the case of the negative control. This fact is due to, in a transient expression system, the protein is highly overexpressed and a great amount of protein accumulates in nucleolus producing unspecific interaction in this part of the nucleus. So, the experiment seems to work.

Next, we used BIFC analyses to investigate ING4 dimerization in living cells and for that we examined complementation between VN173N-ING4 and VC155N-ING4 and, between VN173N-ING4 and VC155C-ING4. Expression of either combination of proteins produced uniform fluorescence in the nucleus of the cells, but this fluorescence was brighter in the first combination (Fig 2, D and E). This fact could indicate that the interaction observed in the first combination should be specific.

To confirm that, we used ING4(Δ1-118) a construct where the N-terminal domain, and thus the dimerization site, had been deleted). This mutant protein was fused to the fluorescent protein fragments in a manner identical to the wild-type protein to work as a negative control. In this case, we examined the complementation between VN173N-ING4(Δ1-118) and VC155N-ING4(Δ1-118) with the result that, contrary to what we expected, this combination produced a brighter fluorescence than in the assay with ING4 wild-type fusions (Fig 2, F). This observation indicated that the interaction was not specific, but it is still unclear why in some combinations the fluorescence was more intense than in others. In order to avoid biases due to naked eye evaluation of the images, the different assays, were quantified and the

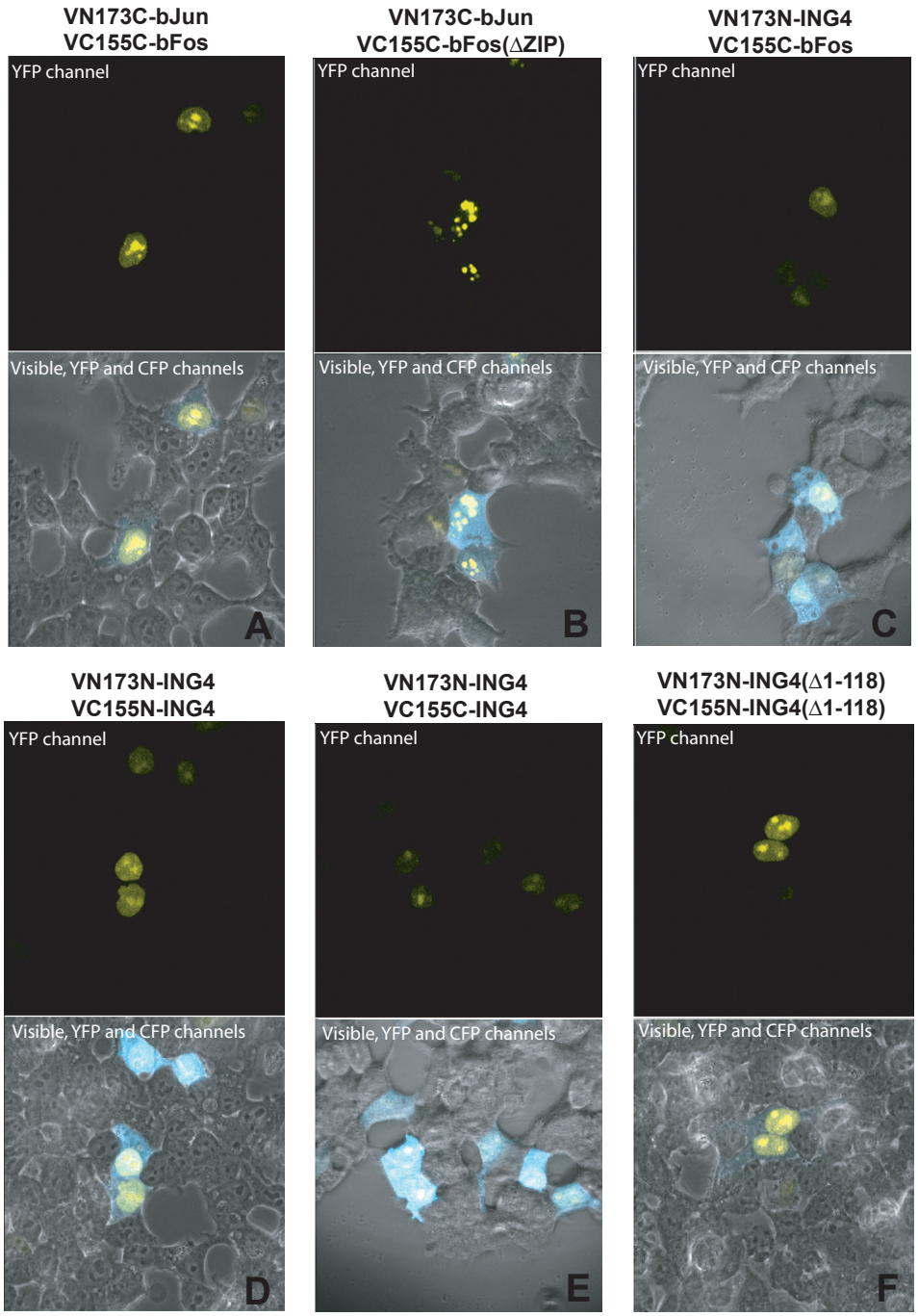

Figure 2. Visualization of ING4 dimerization in living cells. (A-F) Fluorescence images of 293T cells expressing the proteins indicated in each panel were acquired 24h after transfection. The upper image of every panel corresponds to fluorescence in yellow fluorescent protein channel, and the lower with merged image resultant of addition of YFP, CFP and visible channels.

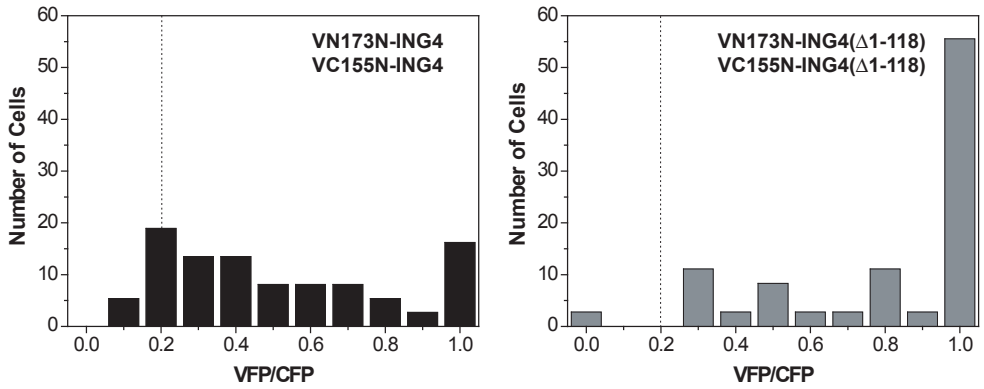

Figure 3. Comparison of the efficiencies of fluorescence complementation by VN173N-ING4 with VC155N-ING4 and VN173N-ING4(Δ1-118) with VC155N-ING4(Δ1-118). The distribution of ratios between VFP and CFP emissions in individual cells cotransfected with plasmids encoding the proteins indicated above each histogram and pECFP is shown.

efficiency of fluorescence complementation by the wild-type and the truncated ING4 was calculated and compared. For this purpose we used a plasmid encoding full-length CFP (cyan fluorescent protein) as an internal reference. In figure 3 we can observe that the complementation efficiency is even larger in the combination of ING4 truncated than in the case of the wild-type.

As an additional negative control to evaluate the specificity of the interaction, we examined the complementation by ING4 (VN173N-ING4) with bFos (VC155C-bFos), which are two proteins have never been describe as interacting partner. Again, contrary to what we expected, this combination also exhibits complementation (Fig 2, C), hence this assay supports the possibility of complementation to be unspecicific.

We can conclude that, al least for ING4, there is fluorescence even in the absence of dimerization mediated by the N-terminal domain. It may be that the BiFC is not an appropriate assay for the study of the ING4 dimerization event, perhaps due to ING4 participating in additional intermolecular interactions, mediated by other regions different from the N-terminal domain. Consequently, this would make possible that two ING4 molecules could remain close in the space. For example, and even, though the N-terminal domain constitutes the dimerization domain *in vitro*, it could be possible that, inside the cells, the macromolecular complexes of chromatin remodeling would have two monomers of ING4 interacting through another component of the complex without the need for N-terminal part of the proteins. However, this could explain the existence of fluorescence even in the absence of dimerization domain but not the fluorescence found in the last negative control, unless the N-terminal region of ING could interact with any protein having the potential form a coiled-coil domain.

References

1. Campos, E. I., Chin, M. Y., Kuo, W. H., and Li, G. (2004) Cell Mol Life Sci 61(19-20), 2597-2613.
2. Gong, W., Suzuki, K., Russell, M., and Riabowol, K. (2005) Int J Biochem Cell Biol 37(5), 1054-1065
3. Shi, X., and Gozani, O. (2005) J Cell Biochem
4. Soliman, M. A., and Riabowol, K. (2007) Trends Biochem Sci 32(11), 509-519
5. He, G. H., Helbing, C. C., Wagner, M. J., Sensen, C. W., and Riabowol, K. (2005) Mol Biol Evol 22(1), 104-116
6. Garkavtsev, I., Kazarov, A., Gudkov, A., and Riabowol, K. (1996) Nat Genet 14(4), 415-420.
7. Bienz, M. (2006) Trends Biochem Sci 31(1), 35-40
8. Mellor, J. (2006) Cell 126(1), 22-24
9. Scott, M., Bonnefin, P., Vieyra, D., Boisvert, F. M., Young, D., Bazett-Jones, D. P., and Riabowol, K. (2001) J Cell Sci 114(Pt 19), 3455-3462
10. Wang, Y., Wang, J., and Li, G. (2006) FEBS Lett 580(16), 3787-3793
11. Gozani, O., Karuman, P., Jones, D. R., Ivanov, D., Cha, J., Lugovskoy, A. A., Baird, C. L., Zhu, H., Field, S. J., Lessnick, S. L., Villasenor, J., Mehrotra, B., Chen, J., Rao, V. R., Brugge, J. S., Ferguson, C. G., Payrastre, B., Myszka, D. G., Cantley, L. C., Wagner, G., Divecha, N., Prestwich, G. D., and Yuan, J. (2003) Cell 114(1), 99-111
12. Jones, D. R., Bultsma, Y., Keune, W. J., Halstead, J. R., Elouarrat, D., Mohammed, S., Heck, A. J., D'Santos, C. S., and Divecha, N. (2006) Mol Cell 23(5), 685-695
13. Kaadige, M. R., and Ayer, D. E. (2006) J Biol Chem 281(39), 28831-28836
14. Feng, X., Hara, Y., and Riabowol, K. (2002) Trends Cell Biol 12(11), 532-538
15. Palacios, A., Garcia, P., Padro, D., Lopez-Hernandez, E., Martin, I., and Blanco, F. J. (2006) FEBS Lett 580(30), 6903-6908
16. Palacios, A., Munoz, I. G., Pantoja-Uceda, D., Marcaida, M. J., Torres, D., Martin-Garcia, J. M., Luque, I., Montoya, G., and Blanco, F. J. (2008) J Biol Chem
17. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. (2006) Nature 442(7098), 100-103
18. Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G., Shi, Y., Cote, J., Chua, K. F., and Gozani, O. (2006) Nature 442(7098), 96-99
19. Kuzmichev, A., Zhang, Y., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002) Mol Cell Biol 22(3), 835-848
20. Garkavtsev, I., Kozin, S. V., Chernova, O., Xu, L., Winkler, F., Brown, E., Barnett, G. H., and Jain, R. K. (2004) Nature 428(6980), 328-332
21. Gunduz, M., Nagatsuka, H., Demircan, K., Gunduz, E., Cengiz, B., Ouchida, M., Tsujigiwa, H., Yamachika, E., Fukushima, K., Beder, L., Hirohata, S., Ninomiya, Y., Nishizaki, K., Shimizu, K., and Nagai, N. (2005) Gene 356, 109-117
22. Ozer, A., Wu, L. C., and Bruick, R. K. (2005) Proc Natl Acad Sci U S A 102(21), 7481-7486

23. Kim, S., Chin, K., Gray, J. W., and Bishop, J. M. (2004) Proc Natl Acad Sci U S A 101(46), 16251-16256
24. Kim, S. (2005) Cell Cycle 4(4), 516-517
25. Doyon, Y., Cayrou, C., Ullah, M., Landry, A. J., Cote, V., Selleck, W., Lane, W. S., Tan, S., Yang, X. J., and Cote, J. (2006) Mol Cell 21(1), 51-64
26. Russell, M., Berardi, P., Gong, W., and Riabowol, K. (2006) Exp Cell Res 312(7), 951-961
27. Zhou, Y., and Grummt, I. (2005) Curr Biol 15(15), 1434-1438
28. Shyu, Y. J., Liu, H., Deng, X., and Hu, C. D. (2006) Biotechniques 40(1), 61-66
29. Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002) Nat Biotechnol 20(1), 87-90
30. Hu, C. D., Grinberg, A. V., and Kerppola, T. K. (2005) Curr Protoc Protein Sci Chapter 19, Unit 19 10
31. Kerppola, T. K. (2008) Methods Cell Biol 85, 431-470
32. Hu, C. D., Chinenov, Y., and Kerppola, T. K. (2002) Mol Cell 9(4), 789-798

METHYLATED HISTONE 3 RECOGNITION BY THE PHD FINGER OF ING4 IN A CROWDED MACROMOLECULAR ENVIRONMENT

Alicia Palacios and Francisco J. Blanco
Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia, 48160 Derio, Spain

Abbreviations: ING, inhibitor of growth; PHD, plant homeodomain; H3K4me3, histone 3 trimethylated at lysine 4; NMR, nuclear magnetic resonance; HSQC, heteronuclear single quantum coherence spectroscopy; CSP, chemical shift perturbation.

The challenge of packaging the chromosomal DNA inside the nuclei of eukaryotic cells is principally met by four conserved histone proteins that assemble into an octamer with the DNA double helix wrapped around it. This structure, the nucleosome, is the basic unit of chromatin, with higher levels of organization achieved with the concourse of additional proteins. The structure of chromatin strongly influences gene expression levels, silencing those genes located in the most compact areas. Chromatin structure regulation is based on reversible covalent histone modifications, the most abundant being methylation and acetylation of lysine residues at the histone N-terminal flexible tails. The specific combinations of these covalent modifications mark the nucleosomes with a histone code that is read by dedicated protein domains [1,2]. These modules are present in proteins that form part of complexes with enzymatic components that add or remove other modifications altering the structure of chromatin with an impact on DNA accessibility and gene transcription. One such reader module is the Plant HomeoDomain, an interleaved finger scaffold consisting of the C4HC3 sequence motif that coordinates two Zn^{2+} ions. The PHD is present in many proteins involved in chromatin remodelling, and specifically recognizes histone-3 trimethylated at lysine 4 (H3K4me3), a hallmark of genes active in transcription [3,4].

The protein ING4 is one of the members of the family tumour suppressors INhibitors of Growth [5]. The five proteins of the family contain a PHD finger at its C-terminal end, and form part of different histone acetylation or deacetylation complexes. The recognition of H3K4me3 by ING4 is based on enthalpic contributions, and a number of hydrogen bonds, ionic pairs and cation-π interactions are identified in the structure of the complex. However the discrimination between the methylated and unmethylated forms of the histone is predominantly of entropic nature, which can be explained by differential solvation effects [6,7]. Interestingly, and in contrast to other PHD fingers, ING4 shows little discrimination between the mono-, di- and trimethylated forms of H3K4. The affinity of the ING4-PHD for the histone peptides has been determined by NMR in defined buffered conditions, which strongly deviates from the native medium regarding the crowded environment of the nucleoplasm, with an extremely high concentration of macromolecules. The effect of this macromolecular crowding on protein folding, stability, and association, has been experimentally addressed by means of highly concentrated solutions of polymers that mimic the crowded interior of cells, and it has been reported that they enhance the binding and the assembly of proteins [8-16]. It is believed that macromolecular crowding enhances these reactions

Table 1.

Dissociation constants (K_D) measured in solution by NMR for H3K4me3/2/1/0 peptides bound to PHD (ING4) in presence of 15 % ficoll

	K_D (0%ficoll)	K_D (15%ficoll)
$H3_{15}K4me3$	3.9 ± 0.7 μM	0.4 ± 0.3 μM
$H3_{15}K4me2$	1.8 ± 1 μM	1.3 ± 0.6 μM
$H3_{15}K4me1$	6 ± 2 μM	12 ± 4 μM
$H3_{15}K4me0$	370 ± 20 μM	100 ± 10 μM

by a non-specific reduction of the total excluded volume [17]. Still, no high resolution structural studies of protein complexes have been done in these media. With the aim of evaluating the macromolecular crowding effects on the binding mode and affinity for ING4 for histone 3 and also its specificity for the different metylation states, we have measured the binding in the presence of Ficoll 70, a commonly used macromolecular crowding agent for *in vitro* studies because it is a sugar polymer that behaves like a semirigid sphere, is chemically inert and does not interact with proteins [18-23]. Preliminar fluorescent measurements at 15% ficoll indicated an increased binding affinity for H3K4me3, but in order to observe the binding mode and the details of the interaction at the residue level and to compare with previous measurements we have used NMR to characterized the binding in the presence of ficoll. This task was not an easy one due to the broadening of the PHD signals, and to the huge intensity of proton signals from the ficoll molecules which were not completely suppressed in the ^{1}H-^{15}N-HSQC spectra of U-^{15}N-PHD finger, reducing the accessible dynamic range, specially in the case of the affinity measurements since t was necessary to measure them at very low protein concentration (10 μM) to obtain dissociation constants precise enough to confidently interpret the observed differences (see supporting information).

The chemical shift perturbations experienced by the residues of the PHD when it binds to histone 3 trymethylated at Lys4 are very similar for all the residues in the absence [6] or in the presence of 15 % ficoll, as can be seen in Fig. 1. Only a few residues show changes barely above the experimental error (supporting figure 1). This result indicates that that the structure of the complex is essentially unchanged by ficoll. The structure of free PHD is also mostly unaffected by ficoll with only a few exposed residues showing CSPs above the error, which could reflect transient and weak binding of ficoll molecules. The measured affinity for the trimethylated histone tail is markedly increased in the presence of ficoll with a linear dependence of the logarithm of the dissociation constant with the concentration of ficoll (figure 2). This enhancement is not uniform for the different methylated states of the histone tails (H3K4me0/1/2/3). While an order of magnitude increase in the affinity is measured for the trimethylated peptide, only a

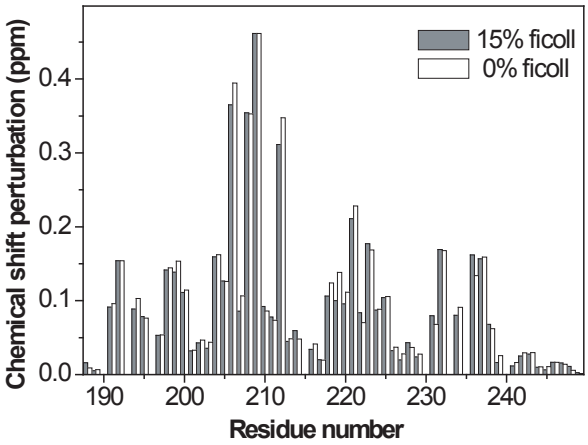

Figure 1. ING4-PHD binding to H3K4me3 in the presence of 15 % ficoll. Binding histogram shows the CSP observed for each residue in the ^{1}H-^{15}N-HSQC spectra of ING4-PHD (25 ºC, 50μM PHD in 20mM sodium phosphate pH 6.5, 50 mM NaCl, 1mM perdeuterated dithiothreitol (d_{10}-DTT), 5% $^{2}H_2O$ and 0.01% NaN_3) in the presence of 1:4 excess of $H3_{15}K4me3$ and with 15% ficoll (white bars) or without it (black bars).

factor of 4 is observed for the unmethylated form, and unchanged values (within the precision of our measurements) for the mono- and dimethylated ones (table 1 and supp. Figure 2). The PHD also sows an increased discrimination more between the trimethylated state of the histone and the other methylated forms in presence of ficoll, therefore we propose that the recognition in a macromolecular environment like the nucleoplam is more specific towards trimethylated form of histone 3 than previously thought. Since the binding mode of other PHD molecules to methylated histone 3 is similar to that of the PHD of ING4, this effect of ficoll could probably be extended to other histone-PHD complexes but this issue needs further measurements. It will be also necessary to do additional experiment to understand the mechanism of the effect of ficoll on the affinity of the PHD for histone 3 tails, in particular the examination of the thermodynamics and the kinetics of the binding and the effect of mixed crowding agents.

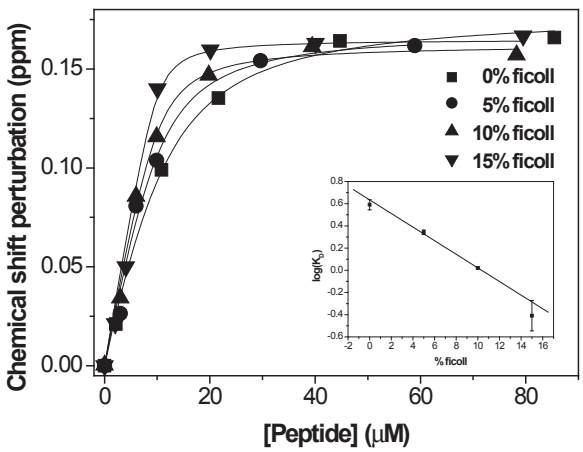

Figure 2. Titrations of ING4-PHD with H3K4meX peptides (X=0,1,2,3) in the presence of different amounts of ficoll. The CSP of W237 amide resonance measured in ^{1}H-^{15}N-HSQC spectra of ING4-PHD is represented against the peptide concentration for each titration. The continuous lines are the fittings to a single site binding model. The height of the symbols corresponds with the experimental error of the measurements. The inset shows the linear correlation between the logarithm of the dissociation constant and the percentage of ficoll. The experimental conditions are as in figure 1 except that the concentration of the proteins was 10 μM.

References.

(1) de la Cruz, X.; Lois, S.; Sanchez-Molina, S.; Martinez-Balbas, M. A. Bioessays 2005, 27, 164-75.

(2) Santos-Rosa, H.; Caldas, C. Eur J Cancer 2005, 41, 2381-402.

(3) Lall, S. Nat Struct Mol Biol 2007, 14, 1110-5.

(4) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. Nat Struct Mol Biol 2007, 14, 1025-40.

(5) Shi, X.; Gozani, O. J Cell Biochem 2005.

(6) Palacios, A.; Garcia, P.; Padro, D.; Lopez-Hernandez, E.; Martin, I.; Blanco, F. J. FEBS Lett 2006, 580, 6903-8.

(7) Palacios, A.; Munoz, I. G.; Pantoja-Uceda, D.; Marcaida, M. J.; Torres, D.; Martin-Garcia, J. M.; Luque, I.; Montoya, G.; Blanco, F. J. J Biol Chem 2008.

(8) del Alamo, M.; Rivas, G.; Mateu, M. G. J Virol 2005, 79, 14271-81.

(9) Gonzalez, J. M.; Jimenez, M.; Velez, M.; Mingorance, J.; Andreu, J. M.; Vicente, M.; Rivas, G. J Biol Chem 2003, 278, 37664-71.

(10) Zhou, Y. L.; Liao, J. M.; Chen, J.; Liang, Y. Int J Biochem Cell Biol 2006, 38, 1986-94.

(11) Hatters, D. M.; Minton, A. P.; Howlett, G. J. J Biol Chem 2002, 277, 7824-30.

(12) Minton, A. P. Curr Opin Struct Biol 2000, 10, 34-9.

(13) Minton, K. W.; Karmin, P.; Hahn, G. M.; Minton, A. P. Proc Natl Acad Sci U S A 1982, 79, 7107-11.

(14) Morar, A. S.; Wang, X.; Pielak, G. J. Biochemistry 2001, 40, 281-5.

(15) Sasahara, K.; McPhie, P.; Minton, A. P. J Mol Biol 2003, 326, 1227-37.

(16) Shtilerman, M. D.; Ding, T. T.; Lansbury, P. T., Jr. Biochemistry 2002, 41, 3855-60.

(17) Zhou, H. X.; Rivas, G.; Minton, A. P. Annu Rev Biophys 2008, 37, 375-97.

(18) Du, F.; Zhou, Z.; Mo, Z. Y.; Shi, J. Z.; Chen, J.; Liang, Y. J Mol Biol 2006, 364, 469-82.

(19) Luby-Phelps, K.; Castle, P. E.; Taylor, D. L.; Lanni, F. Proc Natl Acad Sci U S A 1987, 84, 4910-3.

(20) Monterroso, B.; Minton, A. P. J Biol Chem 2007, 282, 33452-8.

(21) van den Berg, B.; Wain, R.; Dobson, C. M.; Ellis, R. J. Embo J 2000, 19, 3870-5.

(22) Venturoli, D.; Rippe, B. Am J Physiol Renal Physiol 2005, 288, F605-13.

(23) Kozer, N.; Schreiber, G. J Mol Biol 2004, 336, 763-74.

Suplementary Material

NMR sample preparation

The obtaintion of the ING4-PHD, the peptides, the preparation of the NMR samples and the analysis of the chemical shift perturbations on peptide has been previously described in Palacios et al., J Biol Chem 2008.

The stability of the solutions of ING4-PHD/H3K4meX peptides (1:4 molar ratio) in the presence of 15 % ficoll was evaluated under the same conditions used for the NMR measurements monitoring the absorbance of the sample at 280 nm every 8 hours over a period of several days. The absorbance was unchanged during the first 40 hours, with a gradual reduction in the absorbance and the appearance of precipitated material at longer periods of time. Because of this behavior, the NMR titrations were completed in 24 hours or less.

NMR spectroscopy.

In the presence of 15 % ficoll the quality of the HSQC was markedly worse than in its absence due to the increased viscosity of the solution and to the huge intensity of the proton signals from the ficoll molecules, which were not completely suppressed in the ^{1}H-^{15}N-HSQC spectra of the U-^{15}N-PHD finger, reducing the accessible dynamic range. At a concentration of 50 μM PHD, HSQC spectra could be recorded at 800 MHz using a Bruker TCI-cryoprobe in about 2 hours of sufficient quality to measure the chemicals shift perturbations experience by the amide signals (Supp. Figure 3). However, the fitting of these data was very poor as the protein concentration was much larger that the dissociation constant. It was necessary to decrease the PHD concentration but with a signal to noise high enough to measure the chemicals shifts of residue W237 in all the points of the titration. A reasonable compromise between the total duration of the titration and the signal to noise ratio for was achieved using 10 μM PHD solutions and recording HSQC with only 16 t_1 points and a ^{15}N spectral width of 10 ppm. This caused a number of signals to appear folded in the indirect dimension but the carrier was position so that the signal of W237 could be easily identified and not folded (or overlapped with folded signals) over the entire titration. In this way the full titration could be measured under 24 hours distributing the length of each experiment

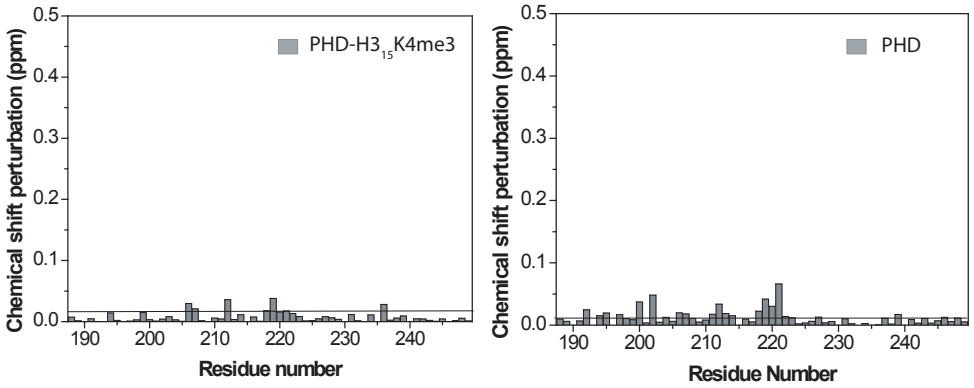

Fig 1. Effect of ficoll on the amide chemical shifts of ING4-PHD free and bound to H3K4me3 peptide. Left: chemical shift perturbations of H3K4me3-bound (left) and free (right) ING4-PHD caused by the presence of 15 % ficoll measured in ^{1}H-^{15}N-HSQC NMR spectra.

to achieve the maximum possible sensitivity (allowing more scans for the mid-transition points since in these spectra the signals were the broadest ones. These measurements could be fitted to a single site binding curve with an error in the fitting that allowed for the observation of significant differences among the peptides and with the concentration of ficoll.

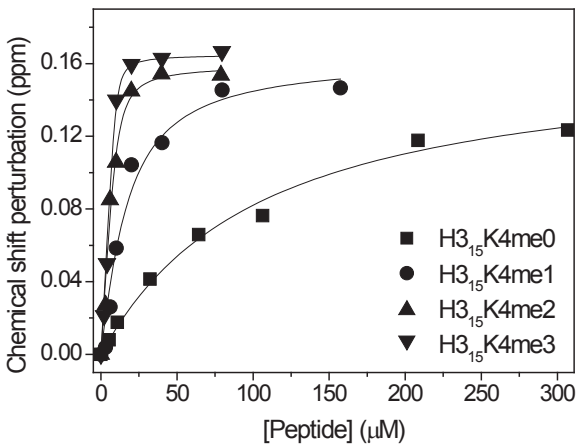

Fig 2. NMR analysis of the binding of PHD to H3$_1$K4meX peptides (3/2/1/0 peptides in presence of 15% ficoll. The CSP of Trp237 amide resonance of PHD in ^{1}H-^{15}N-HSQC spectra is represented as a function of peptide concentration. The continuous lines are fittings to a single site binding model and the height of the symbols corresponds with the experimental error.

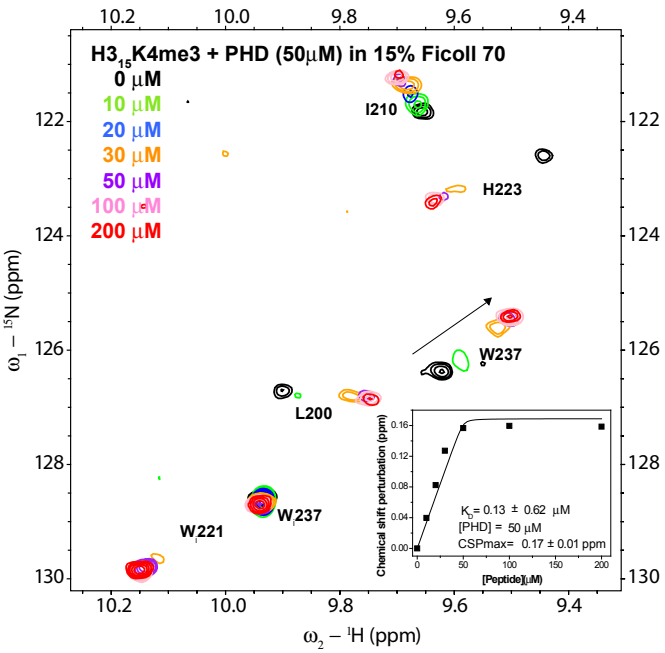

Figure 3. *Left:* Superposition of a region of six ^{1}H-^{15}N HSQC spectra of ING4-PHD after addition of different amounts of H3K4me3 peptide indicated with different colours. The labels adjacent to the signals indicate the corresponding residue ("Wi" stands for the triptophane indol NH signals). Conditions: 25 ºC, 50 μM PHD in 20 mM sodium phosphate pH 6.5, 50 mM NaCl, 1mM perdeuterated dithiothreitol (d$_{10}$-DTT), 5% ^{2}H$_2$O and 0.01% NaN$_3$. *Right:* fitting of the chemical shift perturbations measured in the spectra shown in the left for residue W237 shown the poor fitting obtained in these conditions.

DISCUSIÓN

La caracterización estructural de ING4 y de los dominios individuales definidos de acuerdo a la conservación de su secuencia de aminoácidos y predicciones de estructura confirma que esta proteína consta de tres regiones: el dominio N-terminal (D1) rico en estructura helicoidal, una región central flexible (D2) y el dominio C-terminal (D3) con una estructura de tipo PHD.

Los espectros de RMN y de CD de ING4 y de las distintas construcciones indican que cada dominio es estructuralmente independiente del resto de la molécula, con el dominio flexible y desordenado D2 enlazando los dominios plegados D1 y D3, lo cual les proporciona una movilidad independiente entre sí. ING4 se comporta como una proteína multidominio, donde cada dominio podría tener funciones independientes aunque relacionadas entre sí.

D1 comprende dos motivos de secuencia con importantes implicaciones funcionales: el dominio tipo cremallera de leucina (LZL) y el motivo de reconocimiento de cromatina (PCR). Las medidas experimentales combinadas con la predicción de estructura indican que el dominio D1 podría plegarse en un haz de hélices-α de unos 50 residuos de longitud cada una de ellas, que se corresponden aproximadamente con las secuencias de los dos motivos descritos (LZL y PCR), y que se ensamblan en una estructura terciaria de hélices-α enrolladas (*coiled coil*). Las interacciones entre las hélices pueden ser intra-, inter-moleculares o de ambos tipos. Las medidas de ultracentrifugación analítica demuestran que ING4 se comporta en solución como una proteína dimérica (algo nunca antes descrito experimentalmente para las proteínas ING) siendo el dominio D1 necesario y suficiente para la dimerización de ING4.

La comprobación de que ING4 es también una proteína dimérica en la célula se intentó mediante ensayos de complementación de fluorescencia (BiFC) e inmunoprecipitación. Aunque los experimentos BiFc indican que ING4 dimeriza, no fue posible demostrar la la especificidad de esta unión debido a que el control negativo (la construcción D2D3) dio el mismo resultado. Una posible interpretación podría ser que la ausencia del dominio de dimerización no es un impedimento para que dos monómeros de ING4 se encuentren lo suficientemente cerca en el espacio como para reconstituir la proteína fluorescente fusionada a sus extremos. Esto podría ocurrir si ING4 se encuentra formando parte de grandes complejos junto con otras proteínas, de modo que alguna de ellas podría permitir la cercanía de dos monómeros de ING4 aún en ausencia de D1. Finalmente la dimerización de ING4 dentro de la célula se confirmó mediante ensayos de co-inmunoprecipitación. En sentido estricto la co-inmunoprecipitación de ING4 en células transfectadas con dos construcciones que llevan secuencias reconocidas por anticuerpos diferentes no demuestran que ING4 necesariamente esté formando un dímero, ya que otro tipo de oligómeros y/o la presencia de más de una molécula de ING4 en un complejo sin interacción directa entre ellos producirían el mismo resultado. Sin embargo, teniendo en cuenta que la proteína pura forma dímeros, lo más probable es que este sea también el caso en la célula.

Las medidas de SAXS proporcionan una visión global de la estructura de D1, la cual dispersa los rayos X como una molécula alargada con forma más o menos cilíndrica. El modelo estructural que mejor explica el conjunto de todos los datos experimentales de D1 es el de un plegamiento intramolecular en hélice-bucle-hélice dimerizando en forma de cuatro hélices enrolladas. Sin embargo, el radio del cilindro medido por SAXS es algo mayor que el de un ensamblaje de cuatro hélices, y la longitud del cilindro, es superior al de una hélice individual de 50 residuos. Estas discordancias podrían deberse a distorsiones de la forma cilíndrica, a los residuos flexibles de los extremos de la molécula (y quizás también de los bucles que conectan las hélices), o a una dimerización en la cual las hélices de un monómero no se

encuentran apareadas en toda su extensión con las del otro, sino que uno esté desplazado a lo largo del eje longitudinal de la molécula con respecto al otro. Estos datos no permiten determinar si el ensamblaje de las hélices en el dímero es paralelo o antiparalelo. Tampoco es posible hacer una predicción razonable de cual sería la estructura más estable a partir del análisis de la disposición de los aminoácidos a lo largo de las hélices y de las interacciones inter-hélice resultantes en cada caso.

El dominio N-terminal, además de ser responsable de la dimerización, podría ser también responsable de la interacción de ING4 con alguno de los componentes del complejo HBO1, a través de sus motivos LZL o PCR. Así, ING4 podría interaccionar directamente con la proteína hEAF6 (ortólogo humano del factor 6 asociado a ESA1) del complejo HBO1, la cual también contiene una secuencia con alto potencial para formar estructura de hélices enrolladas. ING4 podría incluso interaccionar con factores de transcripción que también contengan motivos del tipo de cremallera de leucinas, reclutándolos a la cromatina, o secuestrándolos en el citoplasma en caso de las variantes de ING4 sin la región central y que se localizan en el citoplasma.

La región central de ING4 es predominantemente desordenada y flexible, actuando como nexo de unión entre D1 y D3 y como portadora de los motivos de localización nuclear y nucleolar de la proteína. El reconocimiento de estos motivos por parte de los receptores correspondientes está probablemente facilitado por su flexibilidad. El dominio D2 se ha propuesto como responsable de la interacción directa entre ING4 y p53 según lo observado en experimentos de tipo *pull-down* con las proteínas aisladas (aunque usando fusiones de ING4 a GST) (1). La caracterización estructural de esta supuesta unión no ha sido posible ya que no se observa interacción detectable por RMN, lo cual indica que o no hay interacción o es de muy baja afinidad. Este resultado cuestiona los datos obtenidos con los experimentos de tipo *pull down* y apuntan a la posible existencia de otros componentes como mediadores o fijadores de la unión ING4/p53 detectada por otros autores en células mediante ensayos de co-inmunoprecipitación. Funcionalmente se ha propuesto que ING4 y p53 podrían colaborar en la activación de la transcripción dependiente de p53 por dos vías: 1) p53 podría actuar como uno de los factores nucleares que activaran y reclutaran las proteína ING4 a la cromatina (2); 2) ING4, que forma parte del complejo de acetilación de histonas HBO1, podría interaccionar con p53 para facilitar su acetilación y consecuente activación. En ambas vías de colaboración, ING4 y p53 podrían encontrarse formando parte del mismo complejo HBO1, de modo que alguno de los componentes de este complejo podría hacer de nexo de unión entre ambas proteínas (3).

La construcción D3 contiene un dominio PHD de dedos de zinc seguido de una secuencia de aminoácidos básicos. Los espectros de dicroísmo circular muestran que D3 apenas poseía elementos de estructura secundaria regular, lo que concuerda con el plegamiento típico de los dominios PHD (4). La estructura determinada en disolución mediante RMN muestra un plegamiento en el que el motivo de secuencia C4HC3 coordina 2 átomos Zn^{2+}. La secuencia básica en el extremo C-terminal del PHD está desordenada y flexible[1], lo cual podría tener implicaciones en relación con su posible función en el reconocimiento de lípidos de señalización. Así, las proteínas ING1 e ING2 contienen también una región polibásica (PBR) en su extremo C-terminal, que junto con la región carboxilo terminal del PHD, parecen responsables de su unión a fosfatos de fosfatidilinositol (PtdInsPs) (5, 6). La estructura de las regiones

[1]En la estructura del PHD que se describe en el artículo correspondiente al capítulo 2 las regiones N- y C- terminales de la proteína aparecen con estructura definida en la familia de modelos de RMN debido a una interpretación errónea de los NOEs en dichas regiones. Tras comprobar su alto grado de flexibilidad por estudios de dinámica en solución, la estructura fue recalculada y redepositada en el PDB (2K1J).

PBR posteriores a los dominios PHD que unen PtdInsPs se desconoce ya que en todas las estructuras de PHD resueltas la región PBR ha sido delecionada, pero se cree que el PBR junto con algunos residuos del extremo C-terminal del propio PHD forman el sitio de unión a PtdInsPs (5).

El extremo C-terminal de ING4 correspondiente a la región polibásica (PBR) es mucho más corto, contiene sólo cuatro residuos aunque todos son básicos y no reconoce PtdInsPs. Probablemente la flexibilidad de esta región y su pequeño tamaño no permiten el reconocimiento de la cabeza polar de los PtdInsPs. Esto sugiere que ING4 no participa en la respuesta celular mediada por la señalización de PtdInsPs, a diferencia de ING1 e ING2.

El dominio PHD de ING4 se une a la cola N-terminal de histona 3 trimetilada en el residuo de K4 (H3K4me3) al igual que el resto de PHDs de la familia ING) (7), actuando como módulo de lectura del código de histonas. ING4 reconoce con afinidades similares las tres formas metiladas de la histona 3 en K4. Esto le diferencia de los otros PHD de ING1, 2 y 5, que discriminan entre los diferentes estados de metilación (en el caso de ING3 no se ha medido la interacción con H3K4me2/1). ING4 también interacciona, aunque con mucha menor afinidad, con la histona 3 no metilada o metilada en K9 pero, no reconoce la histona 4 (metilada o no en K20), lo que indica su especificidad por la histona 3 metilada en K4.

La estructura del complejo PHD/H3K4me3 muestra como el péptido se acopla en los surcos de la superficie del PHD formando una hebra-β antiparalela, que se aparea con una de las dos hebras-β del PHD. Mediante la formación de esta estructura los residuos R2 (-2 respecto a K4), K4me3 y T6 (+2 respecto a K4) de la histona ocupan los surcos de la superficie del PHD situado en direcciones opuestas con respecto al que ocupa el esqueleto de la cola de histona. El sitio de unión de la lisina metilada está caracterizado por la interacción tipo catión-π entre los dos residuos aromáticos que la rodean y el amonio cuaternario de la lisina metilada. El residuo A7 actúa como bisagra haciendo posible la interacción de los residuos R8, K9 y S10 con el PHD. El extremo N-terminal cargado positivamente se ancla a un bolsillo del PHD, lo que explica la baja afinidad de ING4 por histonas cuyos extremos N-terminales están más alejados de la lisina metilada, como es el caso de H3K9me3. La extensión de la cadena por el extremo C-terminal, sin embargo, no está bloqueada. El complejo se cristalizó usando un péptido correspondiente a los diez residuos N-terminales de la histona 3 ($H3_{10}$K4me3), por lo que cabe la posibilidad de que una extensión del extremo carboxilo del péptido siga interaccionando con el PHD. De hecho, aunque por RMN no se ha llevado a cabo una caracterización exhaustiva del complejo, las medidas ^{1}H-^{15}N HSQC muestran que un grupo de residuos del PHD sufren perturbaciones en sus desplazamientos químicos mayores al unirse a $H3_{15}$K4me3 que a, y estos residuos se localizan en la zona próxima donde se une el extremo C-terminal de $H3_{10}$K4me3.

La mayor afinidad del PHD por los estados de menor metilación de H3K4 en péptidos de quince residuos con respecto a péptidos de diez aminoácidos demuestra la existencia de interacciones histona-proteína más allá del residuo S10, pero sería necesaria resolver la estructura del complejo con un péptido de mayor longitud para saber exactamente cuántos residuos de la cola de histona 3 sería capaz de reconocer el dominio PHD. Esta extensión carboxilo terminal en la interacción hace pensar en la posibilidad de que el dominio PHD actúe no sólo en el reconocimiento de histona 3 metilada en K4 sino que también pueda reconocer modificaciones postraduccionales en otros residuos. El examen de la estructura del complejo en relación a las presencia de otras modificaciones descritas para la histona 3 además de H3K4me3 sugiere que las modificaciones H3T3ph o H3R8me3 podrían interaccionar con ING4,

mientras que no lo harían colas con las modificaciones H3S10ph, H3T11ph, H3K9Ac o H3R2me2a que se perfilan como incompatibles con la estructura del PHD. La hipótesis de la incompatibilidad de la metilación simultánea en R2 y en K4, en cuanto a la unión de la histona a ING4 se ha confirmado experimentalmente, observándose que la metilación de R2 disminuye en 4 veces la afinidad del PHD por la metilación de histona 3 en lisina 4. La relevancia de esta incompatibilidad radica en la mutua exclusión de las modificaciones H3K4me3 y H3R2me2a que se ha descrito recientemente en genes activos en transcripción (8, 9).

Estas modificaciones incompatibles con ING4 podrían funcionar como interruptores en el reconocimiento del código de histonas, como ocurre en HP1 que es reclutada y disociada de la heterocromatina por H3K9me3 y H3S10ph, respectivamente (10, 11).

La formación del complejo está dirigida por una entalpía exotérmica de unión y conlleva una contribución entrópica desfavorable para la unión. Sin embargo, a pesar de ser pequeña en relación a la entalpía, es precisamente la diferente entropía de solvatación la responsable de la discriminación entre las distintas formas metiladas y no metilada.

El PHD posee regiones del esqueleto con distinta dinámica y algunas de las más móviles se vuelven más rígidas al formarse el complejo. Estas zonas del esqueleto se corresponden, en general, con las regiones que participan en la interacción con el fragmento de la histona, sugiriendo que la dinámica del PHD puede contribuir al reconocimiento específico de la histona 3 metilada en K4. Esta observación contrasta con la idea predominante hasta ahora que presentaba al PHD como un dominio de reconocimiento estático (12).

Los dominios PHD de ING2 e ING4 reconocen con una afinidad similar la histona 3 trimetilada en lisina 4, pero el modo de reconocimiento muestra importantes diferencias. Mientras que ING4 reconoce las diferentes formas metiladas con similar afinidad, ING2 discrimina más, interaccionando con mayor afinidad con H3K4me3. Además, a diferencia de ING2, ING4 une, aunque con baja afinidad, la histona 3 sin modificar y la histona 3 metilada en K9. Estas observaciones indican que deben existir interacciones adicionales de ING4 con el resto de la histona, que no tienen lugar en el caso de ING2. Observando ambos complejos podemos ver cómo la unión es muy similar respecto a los seis primeros residuos de la histona 3 y cómo es a partir de este residuo cuando el reconocimiento muestra diferencias. El péptido H3K4me3 sólo interacciona con sus seis residuos N-terminales con ING2 frente a los diez o más residuos que interaccionan con ING4. Estas diferencias, inesperadas dada la similitud estructural entre ambos PHDs, son consistentes con el diferente efecto sobre la afinidad de ambos PHDs por H3K4me3 cuando la tirosina de la jaula de aromáticos, un residuo conservado en ambos, se muta a alanina. La respuesta al porqué de esta discrepancia parece encontrarse en la diferente distribución de cargas electrostáticas en ambos PHDs. Flanqueando la jaula aromática y en concreto el residuo de tirosina, existe una región con carga negativa en el caso de ING4 y positiva en ING2. La presencia del anillo de tirosina podría apantallar la carga positiva al otro lado en el complejo con ING2 y su ausencia en el mutante desestabilizaría la unión. Esta diferente carga probablemente podría también contribuir a la diferencia en la interacción con los residuos 8 a 10 de la histona 3. El diferente comportamiento de ING4 e ING2 está también de acuerdo con sus diferentes papeles en remodelación de la cromatina, ya que mientras que ING2 recluta el complejo hNuA4/Tip60 (HDAC) a la cromatina, ING4 recluta el complejo HBO1 (HAT). El reconocimiento de H3K4me3 por los dominios PHD de ING1 e ING5 descrito recientemente,

muestra que, en ambos casos, el modo de unión es el mismo que el previamente descrito para ING2 (7, 13).

Los estudios realizados con el dominio PHD de ING4 aislado parten de la hipótesis de que su comportamiento es el mismo tanto aislado (una molécula de 7 kDa) como formando parte de la proteína completa (un homodímero de 52 kDa), lo cual es razonable a la vista de los espectros de RMN. Sin embargo, el hecho de que ING4 sea un homodímero plantea además el interrogante de si esta circunstancia influye en el reconocimiento de la histona 3 metilada por parte de sus dominios PHD.

Según se ha podido comprobar mediante RMN, el modo y afinidad de la interacción PHD-H3K4me3 no varía cuando el PHD se encuentra aislado o formando parte de la proteína completa, y puesto que ING4 es dimérica, esto muestra que ING4 es una proteína bivalente de unión a histonas. La existencia de dos sitios de unión independientes en el dímero de ING4, puede aumentar la avidez de ING4 (y por tanto del complejo HBO1) por las regiones de la cromatina ricas en dicha modificación, de una forma análoga a como la densidad de integrinas o de sus ligandos en las superficies de la membrana regula la adhesión

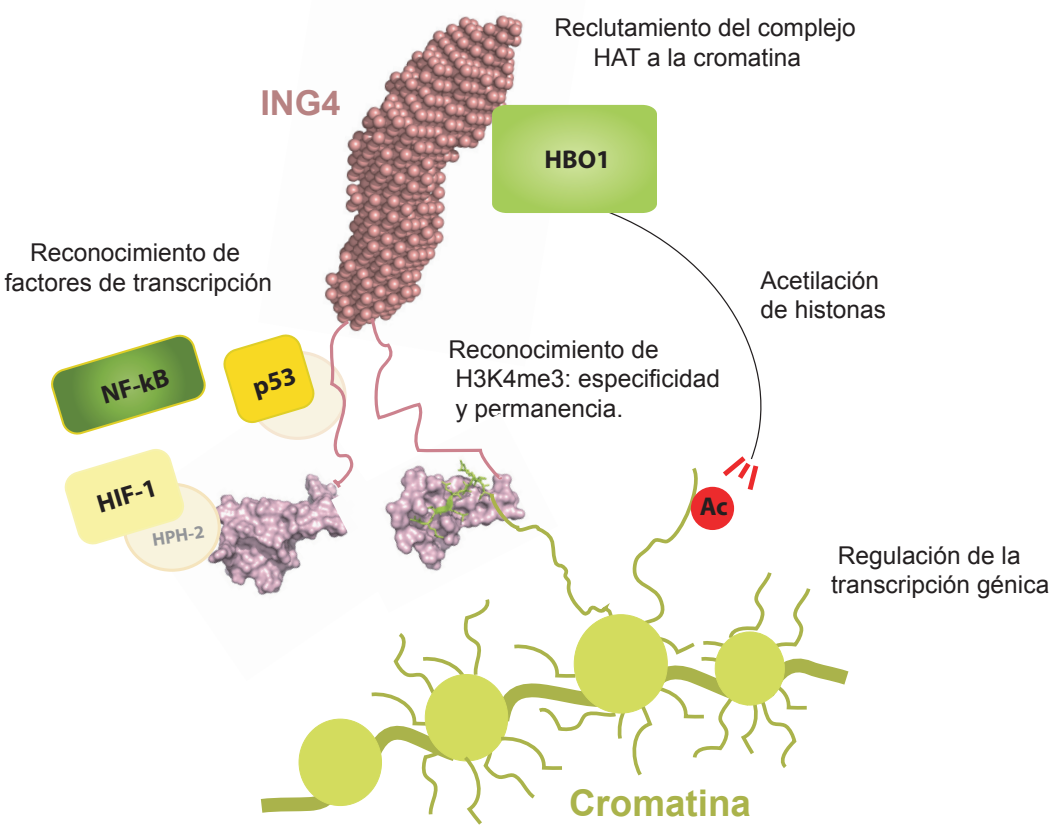

Fig 1. Modelo de las funciones que ING4 y sus dominios pueden tener en la regulación de la transcripción génica a través del reconocimiento de histonas, reclutamiento de complejos HAT e interacción directa o indirecta con factores de transcripción (p53, NF-kB e HIF-1α). El modelo de la estructura de ING4 (en lila) se compone del dominio N-terminal de dimerización y reclutamiento de HBO1 (modelo obtenido por SAXS), el dominio flexible central y el dominio PHD (estructura del dominio PHD aislado resuelta por RMN, a la izquierda, y estructura del complejo con H3K4me3 resuelta por cristalografía de rayos X, a la derecha) de reconocimiento de colas de histonas.

celular (14). En el núcleo, la posible existencia de dos colas de histona 3 de nucleosomas cercanos que puedan unirse a un mismo dímero de ING4 podría también aumentar la eficiencia del reclutamiento del complejo HBO1 a la cromatina con la consiguiente repercusión en el control de la transcripción.

Por otro lado, el entorno molecular en el interior de la célula está siendo reconocido cada vez más como una característica muy importante en las interacciones proteicas, ya que un ambiente con una alta concentración de macromoléculas provoca que el volumen de exclusión (volumen de la disolución no ocupado por las macromoléculas) sea muy pequeño en relación al ocupado, lo que a su vez puede afectar tanto a la estructura de las proteínas y sus ligandos como, a la cinética de la unión y al balance de contribuciones entrópicas y entálpicas de la interacción. Algunos de estos efectos son posiblemente los responsables del aumento de la afinidad de ING4 por H3K4me3 en un orden de magnitud medido en un medio con alta ocupación macromolecular que mimetiza el interior celular. Este hecho muestra la complejidad del sistema y hace plantearse que el funcionamiento del complejo con la histona *in vivo* en cuanto a discriminación entre histonas, permanencia de la interacción o posible unión a más de una modificación o histona es una cuestión aún por responder que aportará nuevas claves en el entendimiento del código de histonas.

En un posible modelo de actuación de ING4, la proteína dimérica formada por dominios estructuralmente independientes sería transportada al núcleo, donde el dominio PHD podría interaccionar con la cromatina en sitios específicos a través de marcas en las histonas (H3K4 metilada) y el dominio N-terminal de ING4 reclutaría hacia la cromatina al complejo de acetilación de histonas HBO1. Este doble reconocimiento promovería la continuidad en el tiempo y especificidad de la unión de ING4 a la cromatina, y facilitaría la actuación de HBO1 sobre la cromatina con la consiguiente represión o activación transcripcional[2] de determinados genes probablemente implicados en la supresión tumoral (figura 1).

ING4 activa el promotor de $p21^{WAF1}$ lo cual constituye un ejemplo de activación de la transcripción. El incremento de los niveles de la proteína $p21^{WAF1}$ (consecuencia de la activación del promotor endógeno) y el bajo nivel de activación cuando el promotor es exógeno (como parte de un vector, sin estructura de cromatina) indica que la activación está directamente relacionada con la regulación de la estructura de la cromatina. Así ING4 podría reclutar el complejo HBO1 (u otros complejos aún por identificar) a la región de la cromatina del promotor de $p21^{WAF1}$ y provocar así la remodelación de la cromatina para que pueda actuar p53 (el factor de transcripción que regula directamente la expresión de $p21^{WAF1}$). Este sería uno de los posibles mecanismos de cooperación entre ING4 y p53. Por otro lado, ING4 interviene en la represión de factores proangiogénicos cuya transcripción está regulada por NF-κB. Sin embargo, en este caso no parece relacionado con el remodelado de cromatina, sino con la inhibición de NF-kB a través de la interacción directa con p65 (uno de los componentes de NF-kB).

Además, muy recientemente se ha descrito que ING4 aumenta el nivel de transcripción de HPH-2, con lo que ING4 no sólo inhibe la ruta de señalización de angiogénesis mediada por HIF a través de la interacción con HPH-2, sino que también lo hace mediante la regulación de la transcripción de HPH-2 (15).

[2]En la discusión de los capítulos 4 y 5 se implica a ING4 en activación de la transcripción debido al reclutamiento de complejos HAT (involucrados en activación de la transcripción). Sin embargo el complejo HBO1 parece ser una excepción dentro de los complejos HAT pudiendo actuar como represor y activador transcripcional.

Los factores de transcripción con los que directa o indirectamente interacciona ING4 (p53, NF-κB, HIF-1α) también podrían actuar como marcas adicionales en la cromatina para el reclutamiento de complejos HAT por parte de de ING4.

La diferente modulación de la transcripción en función de las marcas de la cromatina y las propiedades tanto de activación como de represión de la transcripción del complejo HBO1 (posiblemente en función de sus componentes), junto a que la regulación de la transcripción afecta a genes con efectos finales muy diversos, explica que ING4 pueda estar implicada en muy diferentes funciones sin vínculo aparente entre sí. El panorama funcional se amplía aún más, si tenemos en cuenta que ING4 y sus variantes pueden localizarse en el citoplasma con interacciones que están aún por caracterizar en detalle.

Bibliografía

1. Zhang, X., Wang, K. S., Wang, Z. Q., Xu, L. S., Wang, Q. W., Chen, F., Wei, D. Z., and Han, Z. G. Nuclear localization signal of ING4 plays a key role in its binding to p53. Biochem Biophys Res Commun, *331:* 1032-1038, 2005.
2. Soliman, M. A. and Riabowol, K. After a decade of study-ING, a PHD for a versatile family of proteins. Trends Biochem Sci, *32:* 509-519, 2007.
3. Shi, X. and Gozani, O. The fellowships of the INGs. J Cell Biochem, 2005.
4. Bienz, M. The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci, *31:* 35-40, 2006.
5. Gozani, O., Karuman, P., Jones, D. R., Ivanov, D., Cha, J., Lugovskoy, A. A., Baird, C. L., Zhu, H., Field, S. J., Lessnick, S. L., Villasenor, J., Mehrotra, B., Chen, J., Rao, V. R., Brugge, J. S., Ferguson, C. G., Payrastre, B., Myszka, D. G., Cantley, L. C., Wagner, G., Divecha, N., Prestwich, G. D., and Yuan, J. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell, *114:* 99-111, 2003.
6. Kaadige, M. R. and Ayer, D. E. The polybasic region that follows the plant homeodomain zinc finger 1 of Pf1 is necessary and sufficient for specific phosphoinositide binding. J Biol Chem, *281:* 28831-28836, 2006.
7. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature, *442:* 100-103, 2006.
8. Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlautz, H., Luscher, B., and Amati, B. Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. Nature, 2007.
9. Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bahler, J., Green, R. D., and Kouzarides, T. Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. Nature, 2007.
10. Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature, *438:* 1116-1122, 2005.
11. Hirota, T., Lipp, J. J., Toh, B. H., and Peters, J. M. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature, *438:* 1176-1180, 2005.
12. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol, *14:* 1025-1040, 2007.
13. Champagne, K. S., Saksouk, N., Pena, P. V., Johnson, K., Ullah, M., Yang, X. J., Cote, J., and Kutateladze, T. G. The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. Proteins, *72:* 1371-1376, 2008.
14. Carman, C. V. and Springer, T. A. Integrin avidity regulation: are changes in affinity and conformation underemphasized? Curr Opin Cell Biol, *15:* 547-556, 2003.

15. Tiffany Hung, O. B., Karen S. Champagne, Alex J. Kuo, Kyle Johnson, Howard Y. Chang, and Matthew D. Simon, T. G. K., and Or Gozani ING4 Mediates Crosstalk between Histone H3 K4 Trimethylation and H3 Acetylation to Attenuate Cellular Transformation. Molecular Cell, 33: 248-256, 2009.

CONCLUSIONES

- **La proteína ING4 está organizada en tres dominios estructuralmente independientes: un dominio N-terminal de estructura hélice-bucle-hélice, un dominio central flexible y sin estructura regular persistente, y el dominio PHD C-terminal, con la estructura globular característica de este tipo de dominios y sin interacción con el resto de la proteína.**

- **ING4 forma dímeros tanto aislada en disolución como en el interior de células vivas. La dimerización se produce a través del dominio N-terminal que forma una estructura de hélices enrolladas compuesta, probablemente, por cuatro hélices individuales.**

- **ING4 no interacciona con p53 cuando ambos están aislados en solución. Para que la interacción descrita previamente tenga lugar es probablemente necesaria la participación de otras biomoléculas.**

- **La corta región polibásica situada en el extremo C-terminal del dominio PHD no es suficiente para la unión a la cabeza polar cargada negativamente de fosfatidil inositol fosfatos. ING4, a diferencia de ING1 e ING2, no reconoce PtdInsPs y por lo tanto no participa en la respuesta celular mediada por la señalización de PtdInsPs.**

- **El dominio PHD de ING4 actúa como un módulo lector del código de histonas reconociendo específicamente la histona 3 metilada en el residuo de lisina 4.**

- **La estructura cristalina del complejo PHD/H3K4me3 muestra una gran complementariedad de forma con el péptido, que se pliega en estructura de hebra-β formando una lámina-β antiparalela con las otras dos hebras del PHD. En la estructura del complejo se observan múltiples interacciones de tipo enlace de hidrógeno, catión-π y electrostáticas.**

- **La formación del complejo está dirigida por una entalpía exotérmica de unión y conlleva una contribución entrópica desfavorable para la unión. Sin embargo, es precisamente la diferente entropía de solvatación la base de la discriminación entre de las distintas formas metiladas.**

- **En el esqueleto del PHD de ING4 existen regiones con distinta dinámica. Algunas de las zonas con mayor flexibilidad participan en el reconocimiento molecular de la histona produciéndose un aumento de la rigidez local del PHD con la formación del complejo.**

- **El modo de reconocimiento de histonas por ING4 es diferente del mostrado por ING2, siendo la principal diferencia que la región de la histona 3 que interacciona con el dominio PHD de ING4 es más larga (tres residuos más hacia el extremo C-terminal) que la que lo hace con ING2. Esto es consistente con las diferencias de afinidad por los péptidos y con su diferente papel en el reclutamiento a la cromatina de complejos de acetilación (ING4) o de desacetilación (ING2) de histonas.**

- **La afinidad de la interacción PHD-H3K4me3 no varía cuando el PHD está aislado o forma parte de la proteína completa, por lo que ING4 debe funcionar como un módulo de unión bivalente a histona 3 metilada en el residuo de lisina 4.**

- **En presencia de macromoléculas que mimetizan la alta ocupación del interior celular, el modo de unión del complejo PHD/H3K4me3 no cambia pero la afinidad de la interacción aumenta en un orden de magnitud, lo que refleja la importancia del entorno macromolecular en este tipo de interacciones.**

ANEXOS

FEBS Letters 580 (2006) 6903–6908

Solution structure and NMR characterization of the binding to methylated histone tails of the plant homeodomain finger of the tumour suppressor ING4

Alicia Palacios, Pascal Garcia, Daniel Padró, Eva López-Hernández, Irene Martín, Francisco J. Blanco*

NMR Group, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández Almagro 3, 28029 Madrid, Spain

Received 1 November 2006; revised 17 November 2006; accepted 19 November 2006

Available online 30 November 2006

Edited by Miguel De la Rosa

Abstract Plant homeodomain (PHD) fingers are frequently present in proteins involved in chromatin remodelling, and some of them bind to histones. The family of proteins inhibitors of growth (ING) contains a PHD finger that bind to histone-3 trimethylated at lysine 4, and those of ING1 and ING2 also act as nuclear phosphoinositide receptors. We have determined the structure of ING4 PHD, and characterised its binding to phosphoinositides and histone methylated tails. In contrast to ING2, ING4 is not a phosphoinositide receptor and binds with similar affinity to the different methylation states of histone-3 at lysine 4.

Keywords: ING4; PHD; Histone binding; Phosphoinositide binding; Tumour suppressor; Chromatin remodelling

1. Introduction

The inhibitor of growth (ING) family of tumour suppressors [1] consists of five homologous proteins which form stable complexes with other proteins involved in the regulation of chromatin acetylation [2]. N-terminal histone tail modification is a key mechanism of regulation of chromatin structure, and the pattern of histone modification around a gene affects its transcription [3]. Histone acetylation and methylation at lysines are the most common modifications, and are recognised by specific protein domains [4]. ING proteins contain a conserved C-terminal plant homeodomain (PHD) finger [5], also present in many nuclear proteins involved in gene expression regulation and chromatin remodelling [6]. The PHD of p300 and ACF1 bind to nucleosome histones, and since both proteins contain also a bromodomain, which recognise acetylated lysines, they could form an integrated nucleosome recognition module [7,8]. PHD fingers could bind preferentially to methylated ones, as do chromodomains. This has been confirmed by the recent report that the PHD of ING proteins [9,10] and the PHD of NURF [11,12] bind to histone-3 trimethylated at lysine 4 (H3K4me3).

The PHD fingers of ING1 and ING2 are also nuclear receptors of phosphoinositides [13]. These phospholipids recruit proteins to the vicinity of the membranes regulating cell survival, growth and proliferation. Their interaction with the PHD could regulate the nuclear response to cellular stress [14].

Here, we describe the solution structure of the PHD finger of ING4 and the characterization of its binding to phosphoinositides and histone methylated tails. The results are compared with those reported for ING2 and their functional implications are discussed.

2. Materials and methods

2.1. Protein expression and purification

The PHD finger of ING4 (residues 188–249 with an extra methionine at the N-terminus) was subcloned into the expression vector pET11d from a plasmid harbouring the synthetic gene of ING4 with codons optimized for expression in *Escherichia coli* (Entelechon GmbH). PHD mutants were constructed with QuickChange (Stratagene).

Proteins were over-expressed in *E. coli* BL21(DE3) cells grown at 37 °C in rich medium supplemented with 50 μM $ZnCl_2$ and harvested after 4 h of induction with 0.5 mM isopropyl-beta-D-thiogalactopiranoside. Labeled proteins were produced in minimal medium with $^{15}NH_4Cl$ and [$^{13}C_6$] glucose. After sonication and ultracentrifugation, proteins were found predominantly in the pellet, solubilised in 6 M urea and refolded by a 1:10 fold dilution into cold 20 mM Tris pH 8.0, 50 μM $ZnCl_2$. Purification by anion-exchange chromatography and gel filtration yielded proteins whose identity was confirmed by mass spectrometry. A small amount of wild-type PHD was purified directly from the soluble fraction yielding identical 1D nuclear magnetic resonance (NMR) spectrum as the refolded protein.

2.2. NMR spectroscopy and structure determination

NMR experiments were recorded on Bruker AVANCE 600 (with cryoprobe) and 700 spectrometers at 298 K in 20 mM sodium phosphate pH 6.5, 50 mM NaCl, 1 mM deuterated dithiothreitol (DTT) and 9% or 100% D_2O. Some samples also contained 0.03% NaN_3. Backbone and sidechain resonance assignment were obtained using a set of triple resonance experiments recorded on a 1.2 mM PHD sample. Chemical shifts were measured relative to internal 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS) for ^1H and calculated for ^{15}N and ^{13}C [15]. Spectra were processed with XWINMR (Bruker) or NMRPipe [16] and analyzed using NMRView [17]. Distance restraints were obtained from 2D-NOESY and 3D-NOESY spectra

*Corresponding author. Fax: +34 912246976.
E-mail address: fblanco@cnio.es (F.J. Blanco).

Abbreviations: HSQC, heteronuclear single quantum coherence; NMR, nuclear magnetic resonance; NOESY, nuclear overhauser enhancement spectroscopy; PHD, plant homeodomain; ING, inhibitor of growth; WT, wild type; DTT, dithiothreitol; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt; PIP5, D-myo-phosphatidylinositol 5-phosphate; UV, ultra violet; CSP, chemical shift perturbation

0014-5793/$32.00
doi:10.1016/j.febslet.2006.11.055

edited in ^{13}C or ^{15}N (120 ms mixing time). Dihedral angle restraints were obtained from an HNHA spectrum and from the backbone chemical shifts using TALOS [18]. Structures were calculated with DYANA [19] and used for NOE assignment in an iterative manner. The structures were refined by energy minimization with AMBER 7.0 [20] (see Supplementary Material for the table with the structure statistics for the ensemble of the 25 refined models). The resonance assignment has been deposited with the BMRB entry 7210. The refined models have been deposited in the Protein Data Bank with the entry 2JMQ. These structures are similar to those deposited as PDB entries 1WEN and 1WEU, which contain long segments with extraneous residues at the chain termini and do not include the last four residues of the ING4 sequence.

2.3. Ligand binding

Methylated histone peptides were purchased from NeoMPS, Strasbourg, and contain an extra tyrosine residue at the C-terminus to measure peptide concentration by ultra violet (UV) absorbance. The sequences of the unmodified histone peptides are NH_2-ARTKQTARKSTGGKAY-COOH (residues 1–15 of histone-3), and NH_2-GGAKRHKVLRDNIQY-COOH (residues 14–27 of histone-4). The residues that were methylated in the different peptides are underlined. Stock peptide solutions (5–6 mM) were prepared in 20 mM sodium phosphate pH 6.5, 50 mM NaCl, and the binding was identified by the perturbation in the chemical shifts observed in ^{1}H–^{15}N-HSQC spectra of 50 μM PHD samples in the absence or presence of a 1:4 excess peptide dialysed simultaneously against the same buffer. Titrations were performed by stepwise addition of peptide stock solutions into 50 μM PHD samples and measuring the changes in the chemical shifts of W237 peak in ^{1}H–^{15}N-HSQC. Dissociation constants (K_D) were determined by data fitting (Origin, Microcal) to the equation: $\Delta\delta = (K_D + [P]+[L] - \mathrm{sqrt}((K_D + [P] + [L])^2 - 4*[P]*[L]))/(2*[P])*\Delta\delta_{max}$, where $[L]$ is the concentration of the peptide, $[P]$ is the concentration of PHD, $\Delta\delta$ is the measured chemical shift perturbation (CSP) and $\Delta\delta_{max}$ is the maximum difference in chemical shifts of the free protein and the ligand-bound protein. $\Delta\delta$ was calculated from the equation: $\Delta\delta = \mathrm{sqrt}(((\Delta\delta_H)^2 + (\Delta\delta_N/5)^2)*0.5)$, where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift changes in the ^{1}H and ^{15}N resonances, respectively, upon peptide addition.

Soluble PI5P was from Echelon Biosciences. Its binding was tested with ^{1}H–^{15}N HSQC spectra of 60 μM PHD in the absence or presence of a 1:10 excess of PI5P.

3. Results and discussion

The PHD finger of ING4 requires Zn^{2+} for proper folding. An interleaved finger scaffold consisting of the C4HC3 sequence motif coordinates two Zn^{2+} atoms that stabilise the loops, the antiparallel β-sheet and the one turn helix (Fig. 1), which form the typical fold of this domain [6,21]. The structure is well defined with higher variability at the chain termini, which is due to increased mobility as shown by heteronuclear ^{1}H–^{15}N NOEs (data not shown). The surface of the molecule has a region with high density of positive charge (Fig. 1C) with a large contribution from the last four residues at the C-terminus (RKKK). This region could be involved in the binding to phosphoinositides, in a similar way as a positively charge region present in ING1 and ING2 after the PHD sequences (see Supplementary Material) is necessary for phosphoinositide binding [13]. Similar findings have been reported for the PHD of Pf1 [22]. However, ING4 PHD finger does not bind to a panel of different phosphoinositides, or does so with an extremely low affinity, undetectable in solution by NMR (Fig. 3), and barely detectable in an overexposed dot blot with immobilized phosphoinositides (see Supplementary Material). Phosphoinositide binding is not a property of all PHD fingers [13], not even of those in ING proteins.

Recently, it has been reported that the ING PHD fingers bind to H3K4me3 [9]. The binding site on ING2 has been mapped by NMR and mutagenesis, and the three-dimensional structure of the complex determined by crystallography [10]. We have confirmed by NMR that ING4 PHD finger binds to H3K4me3 peptide, and mapped the binding site (Figs. 2 and 3). The CSP measured in the presence of 1:4 excess of peptide is represented for each residue in Fig. 3. There are many residues that experience large perturbations in their chemical shifts, indicating a large interaction surface. There is a strong similarity with the pattern of changes experienced by ING2 PHD (see Fig. 2 of Ref. [10]), suggesting that the binding site is similar in both proteins. As shown in Fig. 3 (see also Supplementary Material Fig. 2) the PHD finger of ING4 binds to histone-3 and to its six possible methylated variants at K4 or K9 with the same binding site, but it does not bind to histone-4 or its different methylation states in K20 (only for H4K20me3 a few residues show changes just above the experimental error). The titration curves obtained for histone-3 peptides are shown in Fig. 4, and the calculated dissociation constants are summarized in Table 1. H3K4me3 peptide binds with a $K_D = 4.0 \pm 0.7$ μM, close to the value of 7.9 ± 2 μM measured previously [10] (by fluorescence, and possibly not exactly the same PHD chain length). Table 1 also contains the corresponding dissociation constants reported for ING2 PHD [10] for comparison. Both ING2 and ING4 bind to histone-3 methylated tails and not to histone-4, and bind to H3K4me3 with a similar affinity (considering the estimated errors). But beside these similarities there are remarkable differences. ING4

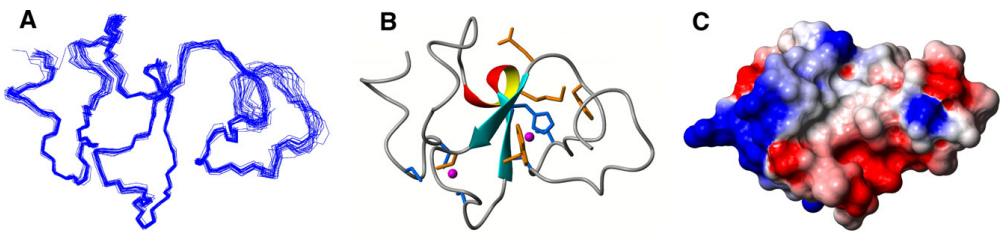

Fig. 1. Solution structure of the ING4 PHD finger: (A) ensemble of 25 refined structures; (B) ribbon model of one of the structures with the two Zn^{2+} ions in magenta, the side chains of the residues coordinating the ions in blue, and the side chains of the residues that experience the largest CSP upon binding to peptide H3K4me3 in orange (Y206, E208, M209, C212, W221, and G235 with CSP above the average plus one standard deviation) and (C) surface of the molecule coloured according to its electrostatic potential (negative in red and positive in blue). The three representations show the molecule in the same orientation.

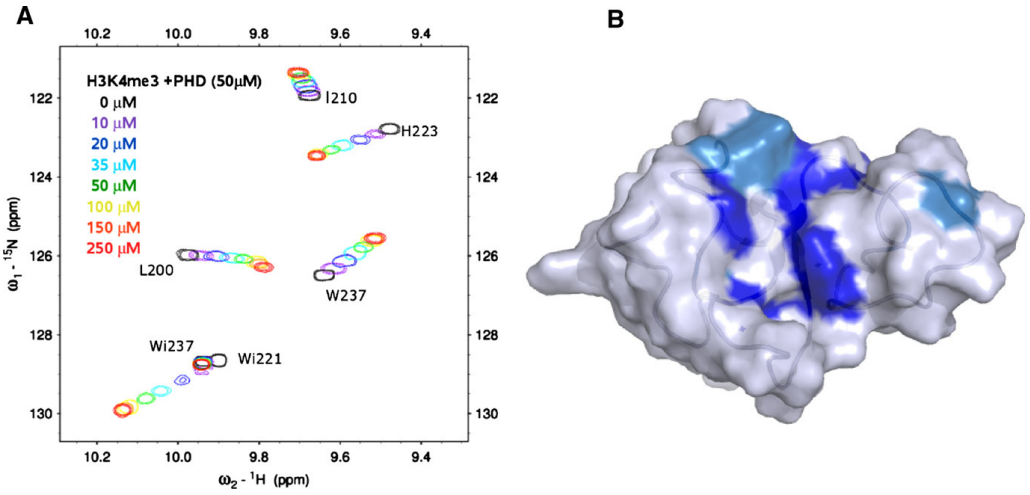

Fig. 2. (A) Superposition of a region of six ^{1}H–^{15}N-HSQC spectra of ING4 PHD after addition of different amounts of H3K4me3 peptide indicated with different colours. The labels adjacent to the signals indicate the corresponding residue (''Wi'' stands for the triptophane indol NH signals). (B) Surface representation of the ING4 PHD finger (in pale blue with a ribbon model inside) with the same orientation as in Fig. 1. Those residues with a CSP upon binding to peptide H3K4me3 larger than the average plus one standard deviation (206, 208, 209, 212, 221, 235) are highlighted in dark blue, and those with a CSP larger than the average (192, 199, 204, 223, 232, 327) in light blue.

PHD is less selective than ING2 towards K4 versus K9 methylated H3. ING4 is also less selective towards the different methylation states of H3K4 than ING2. These measurements suggest a different mode of binding for the PHD. To better characterise the binding site and compare it to the mode of binding of ING2 we have analysed four alanine mutants of ING4 PHD. Y198A mutant corresponds to the Y215A mutation in ING2, which strongly decrease the affinity of ING2 for H3K4me3 [10]. The other three mutations map the relevance for binding of the residue that experiences the largest CSP (M209) and two other residues that are in regions where there is a cluster of large CSP (Fig. 3) and that were not probed in the ING2 study [10]: D192, at the N-terminal region, and W237, which lines a hydrophobic pocket where the methyl groups of the peptide A1 and T2 residues are buried. Mutant W237A is unfolded, as indicated by the sharp and non-dispersed signals observed in its NMR spectra (data not shown), and the addition of H3K4me3 was not enough to displace the folding equilibrium towards an NMR detectable population of the folded state. The other three mutants show dispersed ^{1}H–^{15}N-HSQC spectra and the dissociation constants were measured with a signal that was tentatively assigned to W237 according to its chemical shifts, similar to those measured for the wild-type (WT). As can be seen in Table 2, the three mutants bind to H3K4me3 peptide with affinities that are not very different from those of the WT (reduced or increased by a factor of 3 or 4). These results indicate that M209 plays a minor role in the binding even though is the residue that suffers the largest CSP, probably because it is very close to the trimethylated lysine, as occurs in the ING2 complex [10]. The D192 mutation is in the N-terminal region, and the methylated lysine points towards this end in the structure of the ING2 complex. This explains the observed CSP, and the affinity measured suggests that this region of the protein probably does not make a large contribution to the binding. This residue was not present in the shorter ING2 PHD [10]. Given the similar pattern of the CSP in both PHD fingers, we expected that mutant Y198A would show a very much reduced affinity for H3K4me3, as in ING2. On the contrary, the affinity is reduced but only by a factor of 4 while a three orders of magnitude reduction in the K_D was reported for ING2 PHD [10]. Although the explanation for this difference will only be possible after examination of the structure of the complex of ING4 PHD with the H3K4me3 peptide, this observation demonstrates that there are important differences in the mode of binding for the two molecules. This is consistent with the different selectivity towards the different methylation states of histone-3 discussed above, which is higher for ING2. In this respect, ING4 behaves as the WDR5 module of the MLL1 complex, which activates transcription via methylation of histone-3 [23]. WDR5 binds with similar affinity to the four H3K4 peptides (even the unmethylated one), and is proposed to present the K4 chain of histone-3 for further methylation rather than read its methylation state.

The binding of ING2 PHD to H3K4me3 has been related with transcriptional repression through the recruitment of ING2-HDAC1 complex at target promoters [9,10], and ING2 has been copurified with mSin3A deacetylation complex, which is also linked to repression [2]. However, the PHD finger of NURF helps to recruit the NURF remodelling complex to promoters and modulate transcription initiation [11]. The binding properties of the PHD finger of ING4 link it with actively transcribed genes, since trimethylation of histone-3 at K4 is a hallmark of active genes [24], while trimethylation in K9 (and in K20 of histone-4) is associated with gene silencing [25]. ING4 has been found to copurify with histone-4 acetyl transferase complex HBO1, involved in transcription activation [2]. Although the precise functional implications of

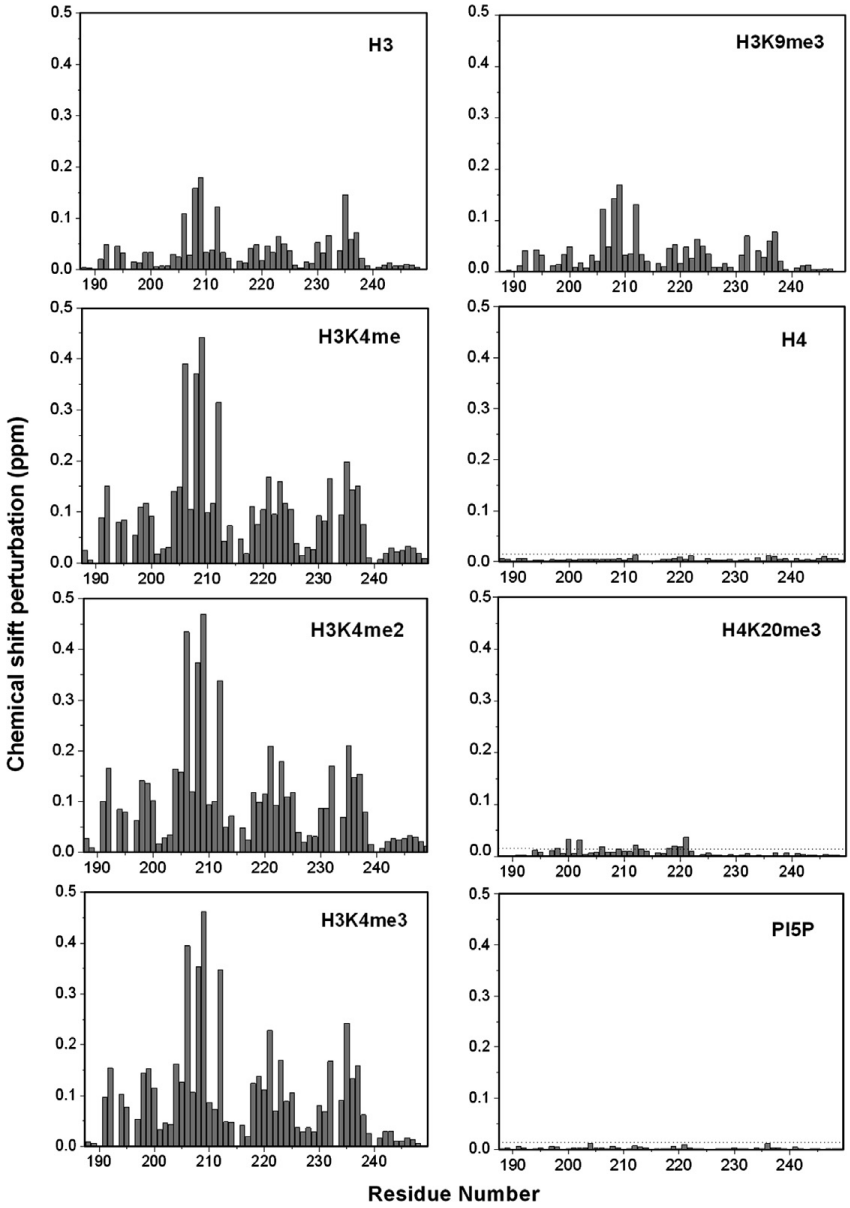

Fig. 3. Binding histograms showing the CSP observed for each residue in the ^{1}H–^{15}N-HSQC spectra of ING4 PHD in the presence of 1:4 excess peptide or 1:10 excess D-myo-phosphatidylinositol 5-phosphate (PIP5). The estimated experimental error (±0.012 ppm) is indicated by the dotted line in the two plots corresponding to peptide H4K20 peptides and PI5P binding.

H3K4me3 recognition by the PHD finger of ING4 are still to be determined, the results presented here show that it behaves differently from ING2. While ING2 PHD performs as a dual specificity module for both H3K4me3 and phosphatidylinositol-5-phosphate [9] and plays a role in transcription repression, ING4 PHD does not bind to phosphoinositides, binds to the three methylation states of H3K4 and is possibly involved in transcription activation.

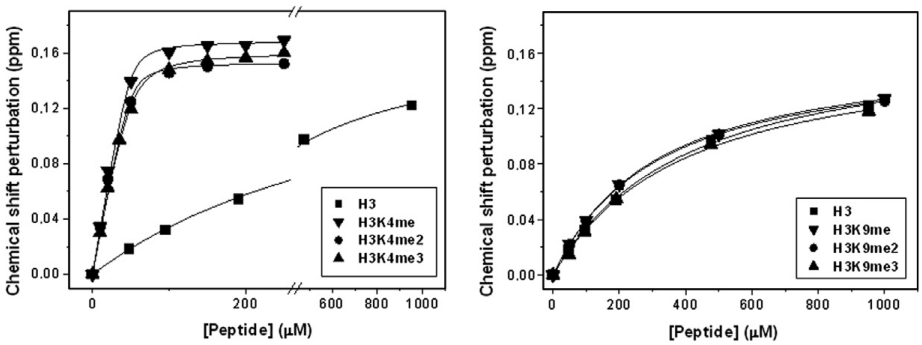

Fig. 4. Plots of the CSP of W237 amide resonance of PHD (50 μM) as a function of the concentration of H3K4 (left), or H3K9 (right) peptides. The symbol's height indicates the experimental error.

Table 1
Dissociation constants for the binding to histone peptides of the PHD fingers of ING4 (this work) and ING2 (Ref. [10])

Ligand	K_D ING4 (μM)	K_D ING2 (μM)
H3	370 ± 20	2240 ± 350
H3K4me	1.6 ± 0.8	208 ± 80
H3K4me2	1.8 ± 1	15 ± 4
H3K4me3	4.0 ± 0.7	1.5 ± 1
H3K9me	274 ± 8	2380 ± 800
H3K9me2	274 ± 6	2320 ± 300
H3K9me3	360 ± 30	2000 ± 60
H4	>10000[a]	n.d.[b]
H4K20me	>10000[a]	>7000
H4K20me2	>10000[a]	>10000
H4K20me3	7900 ± 200[c]	>10000

[a]When no interaction was detected we assume a lower limit of 10000 μM.
[b]Not determined.
[c]Data up to a 1:25 protein:peptide ratio were fitted assuming the same average $\Delta\delta_{max}$ obtained for the other peptides (0.16 ppm).

Table 2
Dissociation constants for the mutants of ING4 PHD (this work) and ING2 (Ref. [10])

Mutant	K_D ING4 (μM)	K_D ING2 (μM)
D192A	0.8 ± 0.6	n.d.[a]
Y198A	12.0 ± 0.8	>5000
M209A	17 ± 4	n.d.[a]
W237A	Unfolded	n.d.[a]

[a]Not determined.

Acknowledgements: We thank David Pantoja and Tahl Zimmerman for help with scripts to work with nmrview and chemical shift data, Jaime Pascual for critical reading of the manuscript and Ramón Campos and Guillermo Montoya for helpful comments. This work was supported by Grants GEN2003-28642 and BIO2003-02246 from the Spanish Ministry of Education and Science (MEC), a grant from Fundación de Investigación Médica Mutua Madrileña, and a MEC Ramón y Cajal contract to F.J.B.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2006.11.055.

References

[1] Gong, W., Suzuki, K., Russell, M. and Riabowol, K. (2005) Function of the ING family of PHD proteins in cancer. Int. J. Biochem. Cell Biol. 37, 1054–1065.
[2] Doyon, Y. et al. (2006) ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol. Cell 21, 51–64.
[3] Santos-Rosa, H. and Caldas, C. (2005) Chromatin modifier enzymes, the histone code and cancer. Eur. J. Cancer 41, 2381–2402.
[4] de la Cruz, X., Lois, S., Sanchez-Molina, S. and Martinez-Balbas, M.A. (2005) Do protein motifs read the histone code? Bioessays 27, 164–175.
[5] Aasland, R., Gibson, T.J. and Stewart, A.F. (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem. Sci. 20, 56–59.
[6] Bienz, M. (2006) The PHD finger, a nuclear protein-interaction domain. Trends Biochem. Sci. 31, 35–40.
[7] Eberharter, A., Vetter, I., Ferreira, R. and Becker, P.B. (2004) ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. EMBO J. 23, 4029–4039.
[8] Ragvin, A. et al. (2004) Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. J. Mol. Biol. 337, 773–788.
[9] Shi, X. et al. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature.
[10] Peña, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R. and Kutateladze, T.G. (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature.
[11] Wysocka, J. et al. (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature.
[12] Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D. and Patel, D.J. (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature.
[13] Gozani, O. et al. (2003) The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell 114, 99–111.
[14] Jones, D.R. et al. (2006) Nuclear PtdIns5P as a transducer of stress signaling: an in vivo role for PIP4Kbeta. Mol. Cell 23, 685–695.
[15] Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J. Biomol. NMR 6, 135–140.
[16] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277–293.
[17] Johnson, B.A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Meth. Mol. Biol. 278, 313–352.

[18] Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR 13, 289–302.
[19] Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. 273, 283–298.
[20] Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E. III, Wang, J., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman P.A. (2002) AMBER 7. University of California, San Francisco.
[21] Pascual, J., Martinez-Yamout, M., Dyson, H.J. and Wright, P.E. (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. J. Mol. Biol. 304, 723–729.
[22] Shi, X. et al. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature 442, 96–99.
[23] Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J. and Verdine, G.L. (2006) Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat. Struct. Mol. Biol. 13, 704–712.
[24] Santos-Rosa, H. et al. (2002) Active genes are tri-methylated at K4 of histone H3. Nature 419, 407–411.
[25] Martin, C. and Zhang, Y. (2005) The diverse functions of histone lysine methylation. Nat. Rev. Mol. Cell Biol. 6, 838–849.

Supplemental Material can be found at:
http://www.jbc.org/cgi/content/full/M710020200/DC1

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 283, NO. 23, pp. 15956–15964, June 6, 2008

Molecular Basis of Histone H3K4me3 Recognition by ING4*[S]

Received for publication, December 10, 2007, and in revised form, March 18, 2008 Published, JBC Papers in Press, April 1, 2008, DOI 10.1074/jbc.M710020200

Alicia Palacios[‡1], Inés G. Muñoz[§1], David Pantoja-Uceda[¶2], María J. Marcaida[§3], Daniel Torres[¶], José M. Martín-García[‖], Irene Luque[‖], Guillermo Montoya[§4], and Francisco J. Blanco[‡5]

From the [‡]Structural Biology Unit, CIC bioGUNE, Parque Tecnológico de Bizkaia, Edificio 800, 48160 Derio, Spain, the Structural Biology and Biocomputing Programme, [¶]NMR Group and [§]Macromolecular Crystallography Group, Centro Nacional de Investigaciones Oncológicas, Melchor Fernández Almagro 3, 28029 Madrid, Spain, and the [‖]Department of Physical Chemistry and Institute of Biotechnology, Faculty of Sciences, University of Granada, 18071 Granada, Spain

The inhibitors of growth (ING) family of tumor suppressors consists of five homologous proteins involved in chromatin remodeling. They form part of different acetylation and deacetylation complexes and are thought to direct them to specific regions of the chromatin, through the recognition of H3K4me3 (trimethylated K4 in the histone 3 tail) by their conserved plant homeodomain (PHD). We have determined the crystal structure of ING4-PHD bound to H3K4me3, which reveals a tight complex stabilized by numerous interactions. NMR shows that there is a reduction in the backbone mobility on the regions of the PHD that participate in the peptide binding, and binding affinities differ depending on histone tail lengths Thermodynamic analysis reveals that the discrimination in favor of methylated lysine is entropy-driven, contrary to what has been described for chromodomains. The molecular basis of H3K4me3 recognition by ING4 differs from that of ING2, which is consistent with their different affinities for methylated histone tails. These differences suggest a distinct role in transcriptional regulation for these two ING family members because of the antagonistic effect of the complexes that they recruit onto chromatin. Our results illustrate the versatility of PHD fingers as readers of the histone code.

Regulation of chromatin dynamics dictates the outcome of fundamental nuclear processes such as DNA transcription replication and repair (1–3). It is central to cell homeostasis, because alterations in chromatin structure contribute to the development of cancer and other human diseases (4). The ING[6] family of tumor suppressors consists of five homologous proteins implicated in chromatin remodeling, growth arrest, and, in cooperation with p53, senescence and apoptosis (5–7). They are frequently deregulated in different types of cancer (8) and contain a conserved C-terminal PHD finger (9) that is present in many nuclear proteins involved in gene expression regulation and chromatin remodeling (10). They form stable histone acetylation or deacetylation complexes (11) and are thought to direct them to specific regions of the chromatin through binding of their PHD fingers to histone 3 N-terminal tails trimethylated at lysine 4 (12, 13). These binding properties link ING proteins with actively transcribed genes, because H3K4 trimethylation is a hallmark of active genes (14). The recognition of H3K4me3 by ING2 is critical for the occupancy of the mSin3A-HDAC1 complex at the promoter of the cyclin D1 gene, which results in histone deacetylation and transcriptional repression of the active gene in response to DNA damage (15). This result suggests a general active transcriptional repression role for ING2; nonetheless, the biological outcome of the recognition of methylated histone tails by the other ING proteins is still unclear. Different PHD fingers link H3K4me3 recognition with gene activation, such as the PHD of the bromodomain PHD finger transcription factor, which helps to recruit the nucleosome remodeling factor complex to target promoters modulating transcription initiation (16, 17). Hence, the function of the PHD-H3K4me3 binding event and its effect on transcription are determined by the particular protein reader of this histone code mark.

We have solved the crystal structure of the PHD of ING4 bound to H3K4me3 tail, which remarkably shows a different mode of binding with respect to the previously reported for ING2, with a longer region of the histone tail participating in the interaction with ING4. The structure of the complex provides new insights into the determinants of the different binding affinities measured for the two domains (13) and allows for the interpretation of NMR data showing the stabilizing effect of C-terminal extensions of the histone peptide and a reduced backbone mobility in the PHD on peptide binding. Isothermal titration calorimetry (ITC) measurements show that the dis-

* This work was supported by a Fundación de Investigación Médica Mutua Madrileña grant and Ikerbasque (Basque Foundation for Science) (to F. J. B.), Ministerio de Educación y Ciencia Grant BFU-2005-02403 (to G. M.), and Ministerio de Educación y Ciencia Grant GEN2003-20642-C09-02 (to G. M. and F. J. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S4 and details on the thermodynamics and dynamics analysis.

The atomic coordinates and structure factors (code 2VNF) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] Both authors contributed equally to this work.
[2] Supported by a Ministerio de Educación y Ciencia Juan de la Cierva contract.
[3] Supported by an EMBO long-term fellowship.
[4] To whom correspondence may be addressed. Tel.: 34-912246983; Fax: 34-912246976; E-mail: gmontoya@cnio.es.
[5] To whom correspondence may be addressed. Tel.: 34-946572521; Fax: 34-946572502; E-mail: fblanco@cicbiogune.es.

[6] The abbreviations used are: ING, inhibitor of growth; PHD, plant homeodomain; H3, histone 3; ITC, isothermal titration calorimetry; ASA, accessible surface area; CSP, chemical shift perturbations; CHD1, chromo-ATPase/helicase-DNA-binding protein 1; HDAC1, histone deacetylase complex 1; NOE, nuclear Overhauser effect; HSQC, heteronuclear single quantum coherence.

TABLE 1
X-ray data collection and refinement statistics

Data collection[a]	
Environment	ADSC detector, ESRF, beamline ID29
Wavelength	1.072 Å
Cell dimensions (Å,°)	$a = 68.51$, $b = 68.51$, $c = 28.51$, $\alpha = \beta = \gamma = 90$
Resolution (Å)	68.51-1.76 (1.82-1.76)
Space group	P43
Unique reflections	12817
Average multiplicity	3.8 (2.6)
Completeness (%)	96.9 (86.8)
R_{merge}[b]	0.075 (0.38)
$<I/\sigma I>$	9.9 (1.7)
Refinement	
Number of reflections (completeness, %)	12178 (96.92)
Resolution range (Å)	68.51–1.76
R factor/R_{free} (%)	15.62/22.67
Number of protein atoms (average B, Å^2)[c]	929 (30.61)
Number of water molecules (average B, Å^2)[c]	129 (38.31)
Root mean square bond length (Å)	0.012
Root mean square bond angle (°)	1.816
Ramachandran plot outliers (number, %)[d]	3, 2.5

[a] The values in the highest resolution shell are given in parentheses.
[b] $R_{merge} = \Sigma_{h\Sigma i}|I_{h,i} - <I_h>|/\Sigma_{h\Sigma i}|I_{h,i}|$.
[c] Calculated using MOLEMAN (43).
[d] Calculated with PROCHECK (44).

crimination in favor of the methylated lysine is due to more favorable solvation entropy contributions.

EXPERIMENTAL PROCEDURES

ING4 Proteins and Histone 3 Peptides Sample Preparation—The clone of ING4 PHD finger 188–246 was made from the 188–249 construct using a QuikChange mutagenesis kit (Stratagene). Protein expression and purification was done as previously described (13). Synthetic lyophilized peptides were purchased from NeoMPS and correspond to histone 3 residues 1–10 (ARTKQTARKS) or residues 1–15 with an extra Tyr residue at the C terminus (ARTKQTARKSTGGKAY) (13) with the four possible methylation states at lysine 4. The concentrations of the peptide stock solutions were measured by amino acid analysis (10 residue peptides) or by ultraviolet absorbance (using the absorbance of the aromatic side chain of the tyrosine residue in the 15 + 1-residue-long peptides).

Crystallization and Structure Determination—Crystals of ING4(188–246) bound to $H3_{10}K4me3$ peptide were grown by the hanging drop vapor diffusion method at 5 °C. The pure protein was dialyzed against 10 mM Tris, pH 6.5, 150 mM NaCl and 2 mM dithiothreitol and concentrated to 1.4 mM. The $H3_{10}K4me3$ peptide was used at a concentration of 35 mM in water. The complex was formed by mixing the PHD and the peptide in a 1:2 molar ratio. Crystallization drops were set up by mixing 1 μl of protein·peptide complex with 1 μl of reservoir solution of 35% polyethylene glycol 6000 and 0.4 μl of 100 mM $CoCl_2$. Needles grew overnight but were not suitable for diffraction experiments. Good quality crystals appeared after using these needles for seeding in fresh drops that were incubated for 3 h at 4 °C. The crystals grew in ~3 days and reached final dimensions of $0.15 \times 0.01 \times 0.01$ mm^3. Prior to data collection, protein complex crystals were immersed in the precipitant solution containing 5% (v/v) ethylene glycol, followed by rapid cooling in liquid nitrogen. A complete x-ray diffraction data set was collected at the beamline ID 29 (European Synchrotron Radiation Facility, Grenoble, France). The data were indexed, integrated, and scaled with HKL2000 (18). The crystals belonged to the P43 space group with cell dimensions $a = b = 68.51$ Å, $c = 28.51$ Å, and $\alpha, \beta, \gamma = 90°$. Matthews coefficient and self-rotation function indicated the presence of two molecules in the asymmetric unit, with a solvent content of 36.78%. Molecular replacement was performed with Phaser (19) using the Protein Data Bank entry 2G6Q as model after removal of the bound peptide. Refinement was carried out with REFMAC5 (20) including rigid body refinement as the first step. Several rounds of iterative rebuilding with O (21) and refinement were performed. Two molecules of 1,4-dithiothreitol and solvent water molecules were placed into the electron density using Arp/Warp (22). The final model contains two complexes in the asymmetric unit: the first was constituted by ING4(195–244)/H3(1–6), and the second was constituted by ING4(191–244)/H3(1–10). Data collection details and statistics of the refinement can be found in Table 1. The coordinates have been deposited with the Protein Data Bank (accession code 2VNF).

NMR Spectroscopy—NMR spectra were recorded at 25 °C in 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, 1 mM dithiothreitol, 5% (v/v) $^2H_2O_2$, 0.01% Na_3N in a Bruker AVANCE 600 as described (13). Titrations were performed by stepwise addition of concentrated (5–6 mM) peptide stock solutions into 600-μl samples of 50 μM PHD. The dissociation constants were determined as described (13). The spectra to measure the differences in the chemical shift perturbation (CSP) upon ING4 binding to $H3_{10}K4m3$ or $H3_{15}K4m3$ were obtained under identical conditions with two samples containing 50 μM PHD and a 4-fold excess of each peptide, which were simultaneously dialyzed against the same buffer. Backbone ^{15}N T_1, T_2, and $\{^1H\}$-^{15}N heteronuclear NOE measurements (23) were performed on a Bruker AVANCE 700 spectrometer on a 0.83 mM uniformly ^{15}N labeled PHD sample with or without 1.66 mM peptide. For free ING4-PHD ten time points (20, 60, 140, 240, 360, 460, 660, 860, 1100, and 1300 ms) were collected for T_1 measurements, and a different set of nine time points (16, 32, 63, 110, 158,190, 222, 270, and 396 ms) was collected to measure the T_2 values. T_1 and T_2 experiments were acquired with eight scans and a rep-

etition delay of 3 s, whereas each $\{^1H\}$-^{15}N NOE spectrum (both saturated and nonsaturated one) was acquired with 136 scans and with an overall recycling delay of 10 s to ensure the maximal development of NOEs before acquisition and to allow solvent relaxation, thus avoiding transfer of saturation to the most exposed amide protons of the protein from scan to scan (24). Relaxation measurements for ING4-PHD bound to H3K4m3 peptides were carried out under the same conditions as those of the free form, but with one more sample time for T_1 experiment (1600 ms) and different number of scans of the $\{^1H\}$-^{15}N NOE experiment 148 and 128 scans in the case of $H3_{10}K4me3$ and $H3_{15}K4me3$, respectively. To check the reproducibility of the relaxation measurements, the T_1 experiments for free ING4-PHD and the $\{^1H\}$-^{15}N NOE experiment for ING4-PHD bound to $H3_{15}K4me3$ were measured twice yielding the same results. The relaxation times were calculated via least squares fitting of peak intensities to a two parameter exponential function, using the rate analysis routine contained in the NMRView program (25). The heteronuclear NOEs were calculated from the ratio of cross-peak intensities in spectra collected with and without amide proton saturation during the recycle delay. Uncertainties in peak heights were determined from the standard deviation of the distribution of intensities in a region of the HSQC spectra where no signal and only noise was observed. The principal components of the inertia tensor were calculated with the Pdbinertia program (A. G. Palmer III, Columbia University) using the first model of the ensemble of 20 structures determined by NMR (Protein Data Bank entry 2JMQ), which is the best one according to restraint violations data and the x-ray structure for the ING4-PHD·H3K4me3 complex. The estimation of the overall correlation time was obtained from the ratio of the mean values of T_1 and T_2, which were calculated from a subset of residues with little internal motion and no significant exchange broadening. This subset excluded those residues with NOEs smaller than 0.65 and also those residues with T_2 smaller than the average minus one standard deviation, unless their corresponding T_1 values were larger than the average plus one standard deviation (26). The diffusion tensor, which describes the rotational diffusion anisotropy, was determined by two approaches (27, 28) using the programs r2r1_diffusion and quadric_diffusion (A. G. Palmer III, Columbia University). The calculations failed when using the errors in T_1 and T_2 estimated by Monte Carlo simulations, which were unrealistically low. Therefore, the errors were scaled up by the minimum factor that allowed an interpretation of the data in terms of a rotational diffusion tensor. This procedure resulted in average errors of 9.5 and 10% for free and bound ING4-PHD, respectively. The ^{15}N relaxation was analyzed assuming dipolar coupling with the directly attached proton, with a bond length of 1.02 Å, and a contribution from the ^{15}N chemical shift anisotropy with a value of −160 ppm. The program FAST-Modelfree (29), which interfaces with the program MODELFREE version 4.2 (30) was used to fit the relaxation data to the model free formalism of Lipari and Szabo (31). Five different models of internal motion were evaluated for each amide ^1H-^{15}N pair: (i) S^2, (ii) S^2 and τ_e, (iii) S^2 and R_{ex}, (iv) S^2, τ_e, and R_{ex}, and (v) S_f^2, S^2, and τ_e, where S^2 is the generalized order parameter, τ_e is the effective internal correlation time, R_{ex} is the exchange contribution to the transverse relaxation, and S_f^2 is related to the amplitude of the fast internal motions.

Isothermal Titration Calorimetry—ITC experiments were performed using a high precision MCS titration calorimetric system (Microcal Inc., Northampton, MA). The ING4-PHD domain was extensively dialyzed against the titration buffer. All of the solutions were filtered, properly degassed to avoid bubble formation, and equilibrated to 25 °C prior to each experiment. The protein solution (at 40–65 μM) in the calorimetric cell was titrated with the appropriate ligand (at 600–800 μM) dissolved in the dialysis buffer following a profile of injection volumes from 2.8 to 20 μl to better define the titration curve. The heat evolved after each peptide injection was obtained from the integral of the calorimetric signal. The heat produced by the binding reaction between the PHD and the peptides was obtained as the difference between the heat of reaction and the corresponding heat of dilution, as obtained from independent titrations of the peptides into the buffer. The resulting binding isotherms were analyzed by nonlinear least square fittings of the experimental data to a model corresponding to a single set of identical sites, as described in the supplemental data. For the interactions of $H3_{10}K4me0$ with ING4-PHD, for which the dissociation constant is out of the range measurable directly by ITC, displacement experiments using $H3_{10}K4me1$ as competing ligand were carried out. Briefly, a 65 μM ING4-PHD solution with $H3_{10}K4me0$ at a 1:4.6 molar ratio was placed in the calorimetric cell and titrated with $H3_{10}K4me1$ (at 2 mM) following a profile of injection volumes from 4 to 20 μl. The resulting binding isotherms, corrected for the dilution heats, were analyzed by nonlinear least square fittings of the experimental data to the exact displacement model as described (32) using the binding affinity and binding enthalpy for $H3_{10}K4me1$ obtained from the titrations with this peptide using the same injection profile. The data analysis was done with Microcal Origin (OriginLab Corporation, Northampton, MA) together with software developed in our laboratory.

Solvation Energy Calculations—The differences in solvation entropy were calculated according to Freire's structural parameterization of the energetics (33) calculated using the crystal structure of the ING4-PHD·$H3_{10}K4me3$ complex and two modeled structures for the di- and mono-methylated species. Of the three methyl groups in the crystal structure, one is fully buried at the binding interface (carbon atom number 891, $\Delta ASA = 0$ Å^2), the second is somewhat exposed (carbon atom number 893, $\Delta ASA = 16.93$ Å^2), and the third one is significantly exposed (carbon atom number 892, $\Delta ASA = 26.93$ Å^2). One could reasonably assume that in the dimethylated complex both methyl groups would tend to be as buried as possible. This is what is observed in the complex of the heterochromatin protein 1 chromodomain with $H3_{15}K9me2$ and me3 peptides, where the methylated lysine side chains adopt a structure very similar to that in ING4-PHD·$H3_{10}K4me3$ complex (34). For this reason, in the model of $H3_{10}K4me2$ bound to ING4 the methyl group that is exposed the most (C atom number 892) was removed from the structure of $H3_{10}K4me3$, whereas C atoms numbers 892 and 893 were removed to build the model for the $H3_{10}K4me1$ complex.

etition delay of 3 s, whereas each $\{^1H\}$-^{15}N NOE spectrum (both saturated and nonsaturated one) was acquired with 136 scans and with an overall recycling delay of 10 s to ensure the maximal development of NOEs before acquisition and to allow solvent relaxation, thus avoiding transfer of saturation to the most exposed amide protons of the protein from scan to scan (24). Relaxation measurements for ING4-PHD bound to H3K4m3 peptides were carried out under the same conditions as those of the free form, but with one more sample time for T_1 experiment (1600 ms) and different number of scans of the $\{^1H\}$-^{15}N NOE experiment 148 and 128 scans in the case of $H3_{10}$K4me3 and $H3_{15}$K4me3, respectively. To check the reproducibility of the relaxation measurements, the T_1 experiments for free ING4-PHD and the $\{^1H\}$-^{15}N NOE experiment for ING4-PHD bound to $H3_{15}$K4me3 were measured twice yielding the same results. The relaxation times were calculated via least squares fitting of peak intensities to a two parameter exponential function, using the rate analysis routine contained in the NMRView program (25). The heteronuclear NOEs were calculated from the ratio of cross-peak intensities in spectra collected with and without amide proton saturation during the recycle delay. Uncertainties in peak heights were determined from the standard deviation of the distribution of intensities in a region of the HSQC spectra where no signal and only noise was observed. The principal components of the inertia tensor were calculated with the Pdbinertia program (A. G. Palmer III, Columbia University) using the first model of the ensemble of 20 structures determined by NMR (Protein Data Bank entry 2JMQ), which is the best one according to restraint violations data and the x-ray structure for the ING4-PHD·H3K4me3 complex. The estimation of the overall correlation time was obtained from the ratio of the mean values of T_1 and T_2, which were calculated from a subset of residues with little internal motion and no significant exchange broadening. This subset excluded those residues with NOEs smaller than 0.65 and also those residues with T_2 smaller than the average minus one standard deviation, unless their corresponding T_1 values were larger than the average plus one standard deviation (26). The diffusion tensor, which describes the rotational diffusion anisotropy, was determined by two approaches (27, 28) using the programs r2r1_diffusion and quadric_diffusion (A. G. Palmer III, Columbia University). The calculations failed when using the errors in T_1 and T_2 estimated by Monte Carlo simulations, which were unrealistically low. Therefore, the errors were scaled up by the minimum factor that allowed an interpretation of the data in terms of a rotational diffusion tensor. This procedure resulted in average errors of 9.5 and 10% for free and bound ING4-PHD, respectively. The ^{15}N relaxation was analyzed assuming dipolar coupling with the directly attached proton, with a bond length of 1.02 Å, and a contribution from the ^{15}N chemical shift anisotropy with a value of −160 ppm. The program FAST-Modelfree (29), which interfaces with the program MODELFREE version 4.2 (30) was used to fit the relaxation data to the model free formalism of Lipari and Szabo (31). Five different models of internal motion were evaluated for each amide ^1H-^{15}N pair: (i) S^2, (ii) S^2 and τ_e, (iii) S^2 and R_{ex}, (iv) S^2, τ_e, and R_{ex}, and (v) S_f^2, S^2, and τ_e, where S^2 is the generalized order parameter, τ_e is the effective internal correlation time, R_{ex} is the exchange contribution to the transverse relaxation, and S_f^2 is related to the amplitude of the fast internal motions.

Isothermal Titration Calorimetry—ITC experiments were performed using a high precision MCS titration calorimetric system (Microcal Inc., Northampton, MA). The ING4-PHD domain was extensively dialyzed against the titration buffer. All of the solutions were filtered, properly degassed to avoid bubble formation, and equilibrated to 25 °C prior to each experiment. The protein solution (at 40 – 65 μM) in the calorimetric cell was titrated with the appropriate ligand (at 600 – 800 μM) dissolved in the dialysis buffer following a profile of injection volumes from 2.8 to 20 μl to better define the titration curve. The heat evolved after each peptide injection was obtained from the integral of the calorimetric signal. The heat produced by the binding reaction between the PHD and the peptides was obtained as the difference between the heat of reaction and the corresponding heat of dilution, as obtained from independent titrations of the peptides into the buffer. The resulting binding isotherms were analyzed by nonlinear least square fittings of the experimental data to a model corresponding to a single set of identical sites, as described in the supplemental data. For the interactions of $H3_{10}$K4me0 with ING4-PHD, for which the dissociation constant is out of the range measurable directly by ITC, displacement experiments using $H3_{10}$K4me1 as competing ligand were carried out. Briefly, a 65 μM ING4-PHD solution with $H3_{10}$K4me0 at a 1:4.6 molar ratio was placed in the calorimetric cell and titrated with $H3_{10}$K4me1 (at 2 mM) following a profile of injection volumes from 4 to 20 μl. The resulting binding isotherms, corrected for the dilution heats, were analyzed by nonlinear least square fittings of the experimental data to the exact displacement model as described (32) using the binding affinity and binding enthalpy for $H3_{10}$K4me1 obtained from the titrations with this peptide using the same injection profile. The data analysis was done with Microcal Origin (OriginLab Corporation, Northampton, MA) together with software developed in our laboratory.

Solvation Energy Calculations—The differences in solvation entropy were calculated according to Freire's structural parameterization of the energetics (33) calculated using the crystal structure of the ING4-PHD·$H3_{10}$K4me3 complex and two modeled structures for the di- and mono-methylated species. Of the three methyl groups in the crystal structure, one is fully buried at the binding interface (carbon atom number 891, $\Delta ASA = 0$ Å^2), the second is somewhat exposed (carbon atom number 893, $\Delta ASA = 16.93$ Å^2), and the third one is significantly exposed (carbon atom number 892, $\Delta ASA = 26.93$ Å^2). One could reasonably assume that in the dimethylated complex both methyl groups would tend to be as buried as possible. This is what is observed in the complex of the heterochromatin protein 1 chromodomain with $H3_{15}$K9me2 and me3 peptides, where the methylated lysine side chains adopt a structure very similar to that in ING4-PHD·$H3_{10}$K4me3 complex (34). For this reason, in the model of $H3_{10}$K4me2 bound to ING4 the methyl group that is exposed the most (C atom number 892) was removed from the structure of $H3_{10}$K4me3, whereas C atoms numbers 892 and 893 were removed to build the model for the $H3_{10}$K4me1 complex.

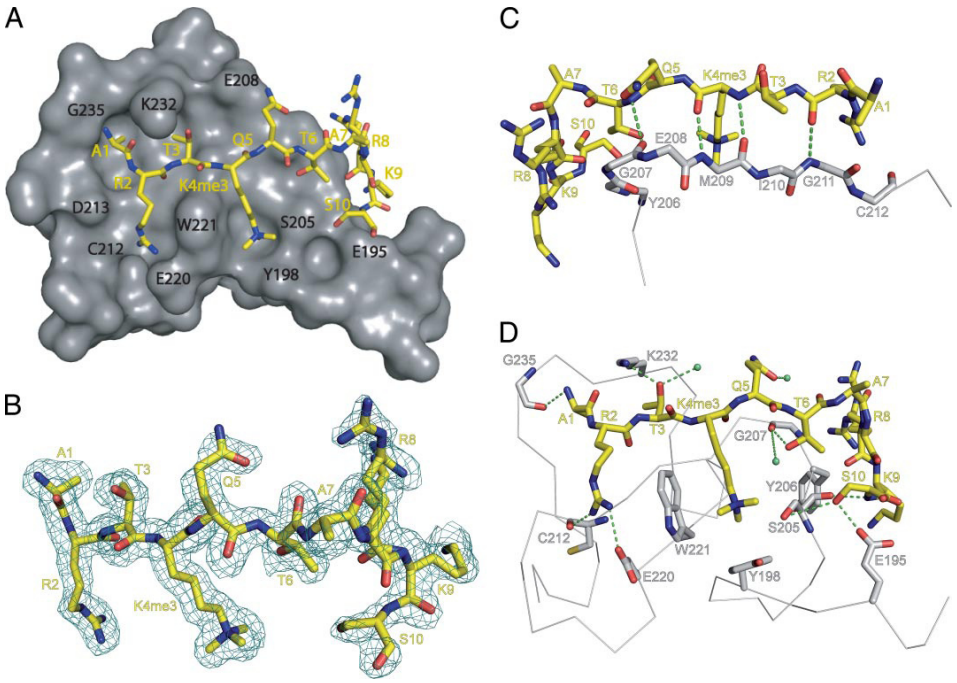

FIGURE 1. **Three-dimensional crystal structure of the complex of ING4-PHD bound to $H3_{10}K4me3$.** *A*, overall structure with ING4-PHD shown in *gray* and $H3_{10}K4me3$ shown in *yellow stick* representation. The location of the ING4 residues lining the binding grooves for histone 3 residues Ala^1, Arg^2, Thr^3, Lys^4, and Ser^{10} is indicated with *black labels*. *B*, view of a $2(|F_o| - |F_c|)$ omit map at 1.76 Å contoured at 1 σ as a *blue mesh*. The omit map was calculated with the program OMIT in the CCP4 package (45). *C*, details of the β-sheet structure formed by H3 residues Arg^2–Thr^6 and ING4 residues Gly^{207}–Gly^{211} with backbone hydrogen bonds in *green*. *D*, details of side chain interactions with hydrogen bonds and water molecules in *green*.

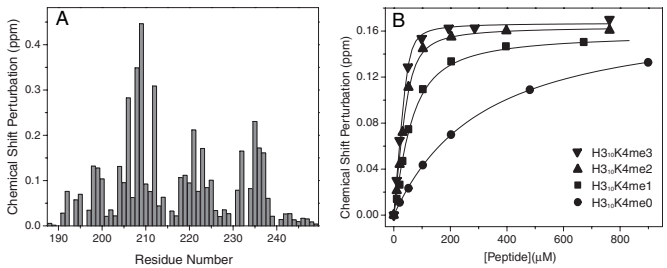

FIGURE 2. **NMR analysis of the binding of ING4-PHD to $H3_{10}K4meX$ peptides.** *A*, binding histogram showing the CSP observed for each residue in the ^1H-^{15}N HSQC spectra of ING4 PHD in the presence of 1:4 excess of $H3_{10}K4me3$. *B*, the CSP of the Trp^{237} amide resonance of ING4(188–249) in ^1H-^{15}N-HSQC spectra is represented as a function of peptide concentration. The *continuous lines* are fit to the binding model (13). The height of the symbol indicates the experimental error.

RESULTS AND DISCUSSION

Recognition of H3K4me3 by ING4—The crystal structure of ING4-PHD (residues 188–246) bound to histone 3 (residues 1–10) trimethylated at lysine 4 ($H3_{10}K4me3$) was solved at 1.76 Å resolution (Fig. 1*A*). The ten residues of the peptide were observed in the electron density, including two alternate conformations for H3 Arg^8 that could be modeled without ambiguity into their corresponding densities (Fig. 1*B*). The peptide conformation does not seem to be influenced by the neighboring crystallographically related molecules, and the interactions with ING4 observed in the crystal structure are consistent with solution NMR data (Fig. 2*A*). The structure of the complex shows that the N-terminal half of the histone tail binds to the surface of the PHD finger as a third strand of the anti-parallel β-sheet that forms the core of the PHD (35). The peptide N terminus forms a hydrogen bond with the carbonyl of ING4 residue Gly^{235} and H3 residues Arg^2–Thr^6, which have β-sheet backbone dihedral angles, form backbone hydrogen bonds with

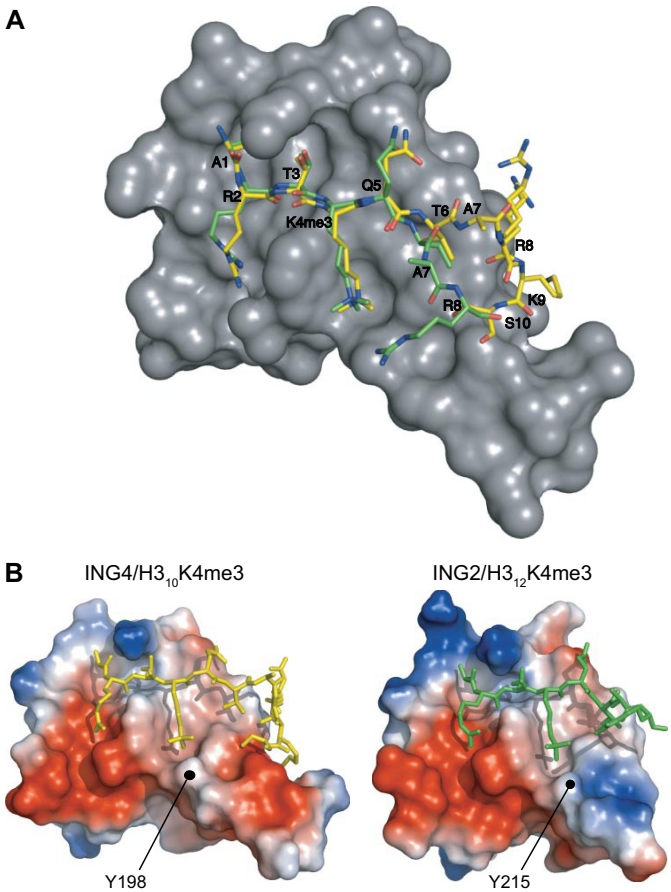

FIGURE 3. **Comparison of the structures of H3K4me3 peptides bound to ING4 and ING2 PHD.** *A*, the crystal structure of ING4-PHD is shown as a *gray surface* with the bound H3$_{10}$K4me3 in *yellow* and the superimposed H3$_{12}$K4me3 from the corresponding ING2-PHD complex structure in *green*. Only H3 residues Ala^{1} to Arg^{8} are seen in the crystal of the ING2 complex (Protein Data Bank entry 2G6Q). *Black labels* indicate the Cα atoms of the peptide residues, whose positions are very similar in both peptides up to H3 Thr^{6}. *B*, surface representations of the PHD fingers of ING2 and ING4 bound to the histone peptides where the electrostatic potential is indicated by a gradient of *red* (negative charge) and *blue* (positive charge) colors. The positions of the homologous tyrosine residues 198 and 215 are indicated in ING4 and ING2, respectively.

Cys^{212}–Gly^{207} (Fig. 1*C*). A kink in the main chain at H3 residues Ala^{7}-Arg^{8} recovers the extended conformation for H3 residues Lys^{9}-Ser^{10}, with H3 Ser^{10} forming a backbone hydrogen bond with Ser^{205}. Because of the polar nature of histone 3 tail residues, the interactions are predominantly polar. The side chain of H3 Arg^{2} forms a salt bridge with Glu^{220}, while the side chains of Thr^{3}, Thr^{6}, and Ser^{10} form hydrogen bonds with Lys^{232}, Gly^{207}, and Glu^{195}, respectively (Fig. 1*D*). The guanidinium group of H3 Arg^{2} and the trimethylammonium group of H3K4me3 make cation-π interactions with the side chain of Trp^{221}, which sits between the two basic peptide side chains. H3K4me3 forms another cation-π contact with Tyr^{198}, which is not essential for recognition because the mutant Y198A still binds the peptide with moderate affinity (13). H3 residue Lys^{9} makes a cation-π interaction with Tyr^{206}, and so does H3 Arg^{8} (at least in one of the two alternative conformations). However, these two interactions are probably weaker because whereas the distances between the charges and the aromatic rings are within the range of typical van der Waals' interactions, their relative orientations are not optimal (the charges are off the C_6 ring axis *C*).

The PHD structure resembles an oblate ellipsoid with three grooves or channels in one side that are filled by the backbone and the side chains of H3 residues Ala^{1}–Thr^{6} (see supplemental Fig. S1). The kink at H3 Ala^{7} directs the peptide chain along the rim of the ellipsoid, and H3 Ser^{10} occupies a small depression lined by Glu^{195}. The shape complementarities between the PHD and the bound peptide structures result in the burial of a large accessible surface area upon complex formation both in the PHD and in the peptide molecules (ASA, 1318 Å^{2}; see below), with H3 Ala^{7} at the kink being the only residue not contributing to the buried ASA.

In the complex with ING4, the methylammonium of H3K4me3 is positioned in a cage formed by two aromatic residues (Tyr^{198} and Trp^{221}), the hydrophobic side chain of Met^{209}, and Ser^{205}. This binding mode of the methylated lysine is common to ING2 and similar to the bromodomain PHD finger transcription factor, with a cage of four aromatic residues, (17). Overall, the recognition of the N-terminal half of H3$_{10}$K4me3 by ING4 is very similar to that observed in the complex with ING2 (12) but differs markedly in the C-terminal half (Fig. 3 and supplemental Fig. S2). The peptide bound to ING2 is not kinked at H3 Ala^{7}, and the last two residues observed in the crystal (H3 Ala^{7}-Arg^{8}) wander off the PHD. The different mode of binding of ING4 is consistent with the different effect of a homologous residue substitution (Y198A in ING4 and Y215A in ING2),

which causes a 3-fold reduction in the affinity of ING4 for $H3_{15}$K4me3 while it strongly destabilizes (more than 3000-fold less affinity) the complex with ING2 (13). The reason for the different effect of the homologous mutation could be the distinct distribution of charges on the surface of the PHD fingers of ING2 and ING4. While in ING4 the N-terminal end of the molecule is predominantly negatively charged on the side closer to K4me3, it is positively charged in ING2 (Fig. 3*B*). The shielding of the aromatic side chain of Tyr^{215} in ING2 may reduce the charge repulsion between this region and K4me3, and its elimination destabilizes the complex beyond the already destabilizing effect of removing the favorable cation-π interaction. In the complex with ING4, the removal of the aromatic side chain of Tyr^{198} eliminates that same cation-π interaction, but it does not result in unfavorable electrostatic interactions.

ING4 Binding to Histone Tails of Different Lengths—The recognition of the H3 Ala^{1} by Gly^{235} blocks N-terminal extensions of the peptide chain bound to ING4, which is probably important for the specificity of the binding of $H3_{10}$K4me3 because the lysine modification is near the N terminus (36). No similar limitation exists at the C terminus, and the extent of the interactions with other regions of the histone tail is unclear.

We have previously characterized the binding of ING4 to 15-residue-long $H3_{15}$K4meX peptides (13). The affinity for $H3_{10}$K4me3 is, within error, the same as for $H3_{15}$K4me3. Nonetheless, whereas ING4 does not discriminate between $H3_{15}$K4me1, -me2, or -me3, it does so for the shorter peptides, with the affinity increasing with the higher number of methyls (Fig. 2*B* and Table 2), indicating that the additional residues contribute to the interaction with ING4. Crystallization trials with this longer peptide were unsuccessful, and the differences in the binding of $H3_{15}$K4me3 and $H3_{10}$K4me3 were examined in solution by NMR measuring the differences in the CSPs caused on the ING4 signals on binding to the two peptides. These measurements are very sensitive to weak interactions or those involving flexible regions of the ligand and/or receptor. The spectra of ING4 bound to $H3_{10}$K4me3 or $H3_{15}$K4me3 (Fig. 4*A*) show several residues with larger perturbations caused by the longer peptide. Moreover, these perturbations are clustered in a region of ING4 in the vicinity of the C-terminal end of the peptide (Fig. 4*B*), indicating that additional interactions take place, resulting in higher affinities for $H3_{15}$K4me1 and -me2 (Table 2). This finding is consistent with the decreased flexibility of the N-terminal end of ING4-PHD bound to $H3_{15}$K4me3 (see below).

Dynamics and Thermodynamics of the Recognition of H3K4me3 by ING4—The chain termini of ING4 are highly flexible in solution, as indicated by the small values of the order parameters (S^2) measured for the backbone ^{15}N atoms (Fig. 5*A*), which are sensitive to movements of the N-H bond in ps to ns time scales. Residue Gly^{235} also shows increased mobility compared with the rest of the chain. Binding to $H3_{10}$K4me3 increases the order parameters at several regions of the ING4

TABLE 2

Dissociation constants (K_D) measured in solution by NMR for the 10- and 15-residue-long histone tails methylated at lysine 4 bound to ING4(188–249)

Ligand[a]	K_D[b]
	μM
$H3_{15}$K4me0	370 ± 20
$H3_{10}$K4me0	274 ± 6
$H3_{15}$K4me1	1.6 ± 0.8
$H3_{10}$K4me1	34 ± 4
$H3_{15}$K4me2	2 ± 1
$H3_{10}$K4me2	9.2 ± 1.4
$H3_{15}$K4me3	3.9 ± 0.7
$H3_{10}$K4me3	3.0 ± 0.6
$H3_{10}$R2me2K4me3	19.2 ± 1.7
$H3_{10}$R2me2[c]	1400 ± 40

[a] The values for $H3_{15}$K4 peptides are from Ref. 13.
[b] The fitting errors are indicated.
[c] Measured from the CSP of C212 signal instead of Trp^{237} (see supplemental Fig. S4).

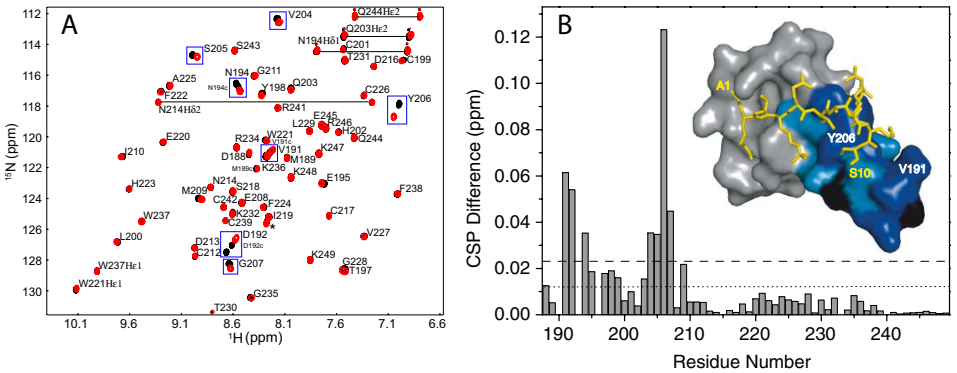

FIGURE 4. **Binding of ING4-PHD to H3K4me3 peptides of different lengths.** *A*, overlay of HSQC spectra of ING4(188–249) bound to $H3_{15}$K4me3 (*black*) or $H3_{10}$K4me3 (*red*). The cross-peaks are labeled with their corresponding residue number and single-letter code. Peaks from amide side chains are connected by *straight lines*. Residues with CSP differences larger than twice the experimental error are *boxed*. Four signals from a minor conformer caused by the *cis-trans* isomerization of the peptidyl-prolyl bonds in the flexible N-terminal region of ING4 are labeled (*M189c*, *V191c*, *D192c*, and *N194c*). A signal from an unknown molecule is marked with an *asterisk*. *B*, binding histogram of the differences in the CSP for each ING4 PHD residue; the experimental error (±0.012 ppm) is indicated by the *dotted line*, while twice this error is indicated by the *dashed line*. The *inset* is a surface representation of ING4 PHD where residues with 0 > ΔCSP > error or ΔCSP > 2 × error are colored in *light* or *dark blue*, respectively. Proline residues for which no signal can be observed are colored in *black*.

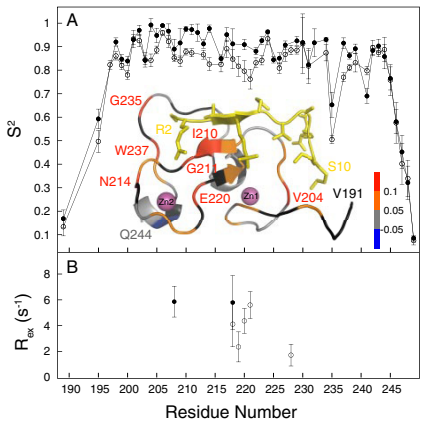

FIGURE 5. **Backbone dynamics of the ING4-PHD backbone.** The backbone ^{15}N nuclei order parameters (*A*, S^2) and the contribution of exchange processes (*B*, R_{ex}) to the relaxation of backbone ^{15}N nuclei are represented *versus* the sequence of ING4, in its free form (*open circles*) and in the complex with $H3_{10}K4me3$ (*filled circles*). The ribbon diagram shows the crystal structure of ING4-PHD bound to $H3_{10}K4me3$ (in *yellow* and with side chains). A color code indicates the change in S^2 for the corresponding ING4 residue on peptide binding. The residues for which the difference could not be measured are colored in *black*.

chain containing residues directly involved in histone recognition (Glu^{195}, Gly^{211}, Glu^{220}, and Gly^{235}). The increased order around Glu^{220} on peptide binding is consistent with the conformational exchange detected for this residue in free ING4 but not in the complex (Fig. 5*B*). Conformational exchange in free ING4 is also detected for Trp^{221}, which interacts with H3 Arg^2 and H3 Lys^4. These results show that the recognition of $H3_{10}K4me3$ is not made by a static PHD molecule but that binding involves regions of ING4 with different degrees of mobility in its free form that become more rigid upon binding to the peptide. Relaxation data on ING4 bound to $H3_{15}K4me3$ show that the flexibility at the N-terminal region is further reduced on binding to this longer peptide (supplemental Fig. S3), which is consistent with the observed contribution to the binding of the longer histone tails. The NMR spectra also show that the free peptides are highly flexible in solution, and conformational order is induced on binding to ING4 (data not shown). A detailed account of the analysis of the backbone ^{15}N relaxation data can be found in the supplemental data.

The energetics of the ING4 binding to $H3_{10}K4meX$ was characterized by isothermal titration calorimetry. Low to moderate binding affinities, with a large increase when the first methyl is introduced, were observed in good agreement with those measured by NMR (Fig. 6). The interaction is driven by a markedly exothermic binding enthalpy partially opposed by unfavorable entropic contributions (Fig. 7). This thermodynamic signature is consistent with the high density of hydrogen bonds and polar interactions at the binding interface and with the burial of apolar and polar ASA upon complex formation (798 and 520 Å^2, respectively, with a ratio $\Delta ASA_{apolar}/\Delta ASA_{polar} = 1.5$). Nonetheless, the methylation state of the lysine residue does not have a significant effect on the enthalpic contributions to the binding affinity, which is very similar for the four peptides, so that the increment in the binding affinity upon methylation is entirely due to changes in the entropic contributions (Fig. 7). It is interesting to note that the opposite situation has been proposed for chromodomains (37). Lysine methylation leads to a polarization of the C_ε-N_ζ bond increasing the cationic character of the methylammonium group and strengthening the cation-π interactions with the aromatic cage but also results in an increment on the hydrophobic character of an otherwise highly polar side chain. The introduction of the methyl groups leads to an increased burial of apolar ASA on complex formation, which results in more favorable solvation entropies. This interpretation is supported by the agreement between the measured differences in the entropic contributions to binding ($-T\Delta\Delta S = -1.9$, -2.8, and -3.3 kcal·mol^{-1} for $H3_{10}K4me1$, -me2, and -me3, respectively, referenced to $H3_{10}K4me0$), and the differences in solvation entropy were calculated (33) in terms of changes in accessible surface area ($-T\Delta\Delta S_{solv} = -1.7$, -2.9, and -2.6 kcal·mol^{-1}).

ING4 Reading of the Histone Code—The structural basis for the binding of ING4 to histone 3 N-terminal tail lies on numerous interactions that occur when the N-terminal histone residues occupy a depression on one side of the PHD. This results, however, in a low binding affinity that increases on methylation at H3 Lys^4. The relevance of the differences in the affinities for mono-, di- or trimethylated peptides is, however, unclear because peptide elongation at the C terminus stabilizes the complexes blurring those differences. This is in contrast with ING2, which does discriminate between the different methylation states of $H3_{12}K4meX$, with ~10-fold affinity increments for every additional methyl (14). The differences in the binding mode between the two ING proteins discussed above may explain these dissimilar affinities, which could be related with a different function. ING2, as part of the mSin3A·HDAC1 complex, links H3K4me3 recognition with transcriptional repression through histone deacetylation. ING4, involved in a complex containing HBO1 (the histone acetyltransferase binding to ORC 1) (11), may link the recognition of H3K4me3 (and of H3K4me2 and -me1) with transcriptional activation through histone acetylation.

The side chain of H3 Arg^2 lies in a groove that is separated by Trp^{221} from another groove occupied by H3K4me3 (Fig. 1*A*), as was described for $H3_{15}K4me3$ bound to the double chromodomain of CHD1 (38). The affinity of the chromodomain binding to H3K4me3 was found to experience a 4-fold reduction when the simultaneous asymmetric dimethylation at H3 Arg^2 (H3R2me2a) occurred on the histone 3 tail. In ING4, the side chain of H3 Arg^2 interacts with the side chain of Glu^{220}, whereas the corresponding position in CHD1 is occupied by a smaller glycine residue. Thus steric clashes may cause an even larger reduction in the affinity of ING4 for K4me3 when the simultaneous H3R2me2a modification occurs on the same histone site. The relevance of this discrimination lies in the mutual exclusion of H3K4me3 and H3R2me2a recently reported in actively transcribed genes (39, 40). The measurement of the binding of ING4 to H3R2me2aK4me3 shows a 6-fold reduction

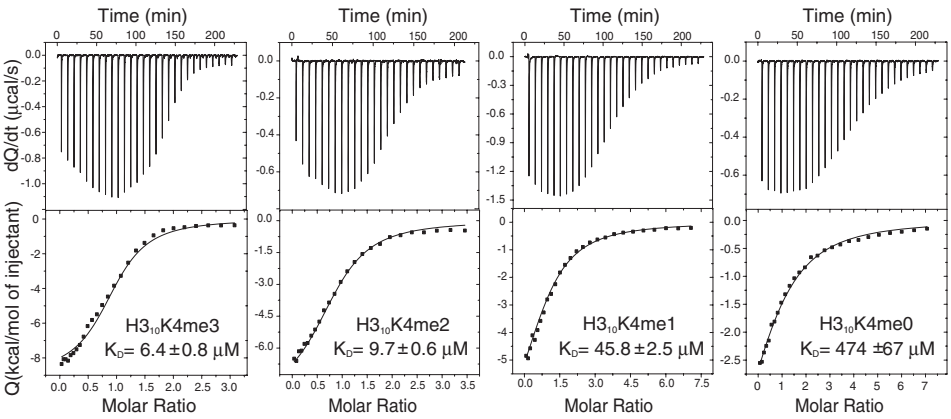

FIGURE 6. **Calorimetric titrations of ING4-PHD with $H3_{10}$K4meX peptides.** Shown are examples of the direct titrations of the $H3_{10}$K4me3, $H3_{10}$K4me2, and $H3_{10}$K4me1 peptides together with the displacement experiment for $H3_{10}$K4me0 using $H3_{10}$K4me1 as competing ligand. In all cases, the *upper panels* represent the heat effect associated with the peptide injections, and the *lower panels* represent the ligand concentration dependence of the heat released upon binding, after normalization and correction for the heats of dilution. The *symbols* correspond to the experimental data, and the continuous *line* corresponds to the best fit to a model of one set of identical binding sites (as described in the supplemental materials) for $H3_{10}$K4me3, -me2, and -me1 and to a complete displacement model for $H3_{10}$K4me0.

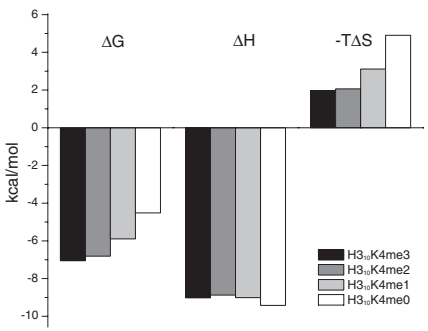

FIGURE 7. **Thermodynamic parameters of the ING4-PHD/H3K4meX binding.** The *bar diagram* shows the change in free energy (ΔG), enthalpy (ΔH), and in the contribution of entropy to the free energy ($-T\Delta S$) of the binding of $H3_{10}$K4meX peptides to ING4(188–249), as determined by ITC. The uncertainty in the experimental values is estimated to be ~5%.

in binding affinity (Table 2 and supplemental Fig. S4), indicating that ING4 can discriminate between chromatin regions enriched in one or both histone modifications. Dimethylation of H3 Arg^2 inhibits the recognition of H3K4me2 by the Spp1-PHD, a subunit of the Set1p histone methylation complex (40), thus regulating trimethylation at H3 Lys^4. Sequence homology suggests that the binding site for H3 Arg^2 in Spp1 is similar to ING4, indicating that dimethylation of H3 Arg^2 will also inhibit binding of ING4 to H3K4me3, which is consistent with the measured reduced affinity for the corresponding histone fragment. In this way, H3R2me2 may help to maintain the chromatin silent not only by inhibiting the recruitment of methyltransferases but also that of histone acetyltransferase complexes such as HBO1. Therefore, understanding how different combinations of histone modifications are read by their recognition modules may be critical to appreciating the regulatory mechanisms exerted through those modifications, as illustrated by the antagonizing role of H3K9me3 and H3S10ph in the recruitment of heterochromatin protein 1 to discrete regions of the chromatin, thereby regulating gene expression (41, 42).

The double chromodomain of CHD1 binds to H3K4me3 with similar affinities when combined with either H3K9me3, H3K9ac, or H3S10ph, but with a 20-fold reduced affinity when H3T3ph is present (38). Based on the structure of the ING4·H3K4me3 complex, phosphorylation of H3T3, a modification correlated with transcription activation, would generate a larger and negatively charged residue that could be accommodated by a conformational change at the side chain of Lys^{232}, with which a favorable electrostatic interaction could be established (Fig. 1*A*). H3K9me3 binds to ING4 with the same affinity as unmodified H3 (13), suggesting the compatibility of methylation at both lysine residues. Acetylation, however, removes the positive charge of H3K9, precluding its interaction with the aromatic side chain of Tyr^{206} (Fig. 1*D*) and probably reducing the affinity of ING4 binding to H3K4me3K9ac. Phosphorylation of H3 Ser^{10} would disrupt the hydrogen bond of its side chain with Glu^{195} (Fig. 1*D*) and introduce an unfavorable electrostatic interaction, suggesting that it would be incompatible with H3K4me3 recognition. A similar effect, though reduced, might occur if H3T11 is the phosphorylated residue because its location on the PHD is close to the negatively charged N-terminal region of the PHD (Figs. 3*B* and 4*B*). On the other hand, monomethylation in Arg^8, which has been associated with repressed transcription, is probably of little relevance for H3K4me3 recognition because its side chain can adopt two

different conformations (Fig. 1), and the structure of the complex does not suggest additional interactions with H3R8me1. Our results suggest that H3K4me3 recognition by ING4 is compatible with H3T3ph, H3R8me1, and H3K9me3, whereas it is not compatible with, or is impaired by, H3R2me2a, H3K9ac, H3S10ph, and H3T11ph. Further experimental studies on PHD fingers and other protein modules will provide more insights into the impact of multiple combinations of histone modifications and their role in the dynamic regulation of chromatin.

REFERENCES

1. Li, B., Carey, M., and Workman, J. L. (2007) *Cell* **128,** 707–719
2. Groth, A., Rocha, W., Verreault, A., and Almouzni, G. (2007) *Cell* **128,** 721–733
3. Downs, J. A., Nussenzweig, M. C., and Nussenzweig, A. (2007) *Nature* **447,** 951–958
4. Rountree, M. R., Bachman, K. E., Herman, J. G., and Baylin, S. B. (2001) *Oncogene* **20,** 3156–3165
5. He, G. H., Helbing, C. C., Wagner, M. J., Sensen, C. W., and Riabowol, K. (2005) *Mol. Biol. Evol.* **22,** 104–116
6. Russell, M., Berardi, P., Gong, W., and Riabowol, K. (2006) *Exp. Cell Res.* **312,** 951–961
7. Soliman, M. A., and Riabowol, K. (2007) *Trends Biochem. Sci* **32,** 509–519
8. Campos, E. I., Chin, M. Y., Kuo, W. H., and Li, G. (2004) *Cell Mol. Life Sci.* **61,** 2597–2613
9. Aasland, R., Gibson, T. J., and Stewart, A. F. (1995) *Trends Biochem. Sci* **20,** 56–59
10. Bienz, M. (2006) *Trends Biochem. Sci.* **31,** 35–40
11. Doyon, Y., Cayrou, C., Ullah, M., Landry, A. J., Cote, V., Selleck, W., Lane, W. S., Tan, S., Yang, X. J., and Cote, J. (2006) *Mol. Cell* **21,** 51–64
12. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O., Zhao, R., and Kutateladze, T. G. (2006) *Nature* **442,** 100–103
13. Palacios, A., Garcia, P., Padro, D., Lopez-Hernandez, E., Martin, I., and Blanco, F. J. (2006) *FEBS Lett.* **580,** 6903–6908
14. Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., and Kouzarides, T. (2002) *Nature* **419,** 407–411
15. Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G., Shi, Y., Cote, J., Chua, K. F., and Gozani, O. (2006) *Nature* **442,** 96–99
16. Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C., and Allis, C. D. (2006) *Nature* **442,** 86–90
17. Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006) *Nature* **442,** 91–95
18. Otwinowski, Z., Minor, W., and Charles, W., Carter, Jr. (1997) *Methods Enzymol.* **276,** 307–326
19. Read, R. J. (2001) *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1373–1382
20. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119
22. Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J., and Perrakis, A. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2222–2229
23. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33,** 5984–6003
24. Renner, C., Schleicher, M., Moroder, L., and Holak, T. A. (2002) *J. Biomol. NMR* **23,** 23–33
25. Johnson, B. A. (2004) *Methods Mol. Biol.* **278,** 313–352
26. Pawley, N. H., Wang, C., Koide, S., and Nicholson, L. K. (2001) *J. Biomol. NMR* **20,** 149–165
27. Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995) *J. Am. Chem. Soc.* **117,** 12562–12566
28. Bruschweiler, R., Liao, X., and Wright, P. E. (1995) *Science* **268,** 886–889
29. Cole, R., and Loria, J. P. (2003) *J. Biomol. NMR* **26,** 203–213
30. Palmer, A. G., Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc.* **113,** 4371–4380
31. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104,** 4559–4570
32. Sigurskjold, B. W. (2000) *Anal. Biochem.* **277,** 260–266
33. Luque, I., and Freire, E. (1998) *Methods Enzymol.* **295,** 100–127
34. Jacobs, S. A., and Khorasanizadeh, S. (2002) *Science* **295,** 2080–2083
35. Pascual, J., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. (2000) *J. Mol. Biol.* **304,** 723–729
36. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007) *Nat. Struct. Mol. Biol.* **14,** 1025–1040
37. Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S., and Waters, M. L. (2007) *Proc. Natl. Acad. Sci. U. S. A.* **104,** 11184–11188
38. Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005) *Nature* **438,** 1181–1185
39. Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlautz, H., Luscher, B., and Amati, B. (2007) *Nature* **449,** 933–937
40. Kirmizis, A., Santos-Rosa, H., Penkett, C. J., Singer, M. A., Vermeulen, M., Mann, M., Bahler, J., Green, R. D., and Kouzarides, T. (2007) *Nature* **449,** 928–932
41. Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. (2005) *Nature* **438,** 1116–1122
42. Hirota, T., Lipp, J. J., Toh, B. H., and Peters, J. M. (2005) *Nature* **438,** 1176–1180
43. Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M., and Jones, T. A. (2001) *International Tables for Crystallography*, Vol. F, pp. 353–367, Kluwer Academic Publishers, Dordrecht, The Netherlands
44. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
45. Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58,** 1955–1957